

When a vaccine is safe

Unfounded public fears place pressures on vaccine developers that go beyond reasonable safety considerations, as the search for an acceptable vaccine against Lyme disease may demonstrate.

Nothing gets forgotten as quickly as the threat of an epidemic that has been successfully headed off. Just look at all the allegations of media hype over SARS (severe acute respiratory syndrome), by people who think it was not a threat because so few ended up contracting it.

The short public memory reflects a low level of awareness of the largely invisible public-health activities that can now prevent infectious diseases running unchecked through human populations like they used to do — and as SARS might have done without the stringent quarantines and travel bans imposed to contain it. In wealthy countries, a lack of personal experience of infectious diseases has also induced a lack of respect for two of the main weapons that keep them at bay — antibiotics and vaccines.

Inappropriate use of antibiotics has allowed microbes to become resistant to many common treatments, creating serious health-care problems. And public suspicion of vaccines can undermine programmes aimed at eradicating particular diseases: the hostility in Britain to the measles, mumps and rubella (MMR) vaccine, based on unfounded fears that it might cause autism, is a case in point (see *Nature* 439, 1–2; 2006).

Vaccination programmes often face problems of public acceptance as, by definition, they treat large numbers of healthy people. It is easier to provide a convincing case for vaccination when the risk of catching a disease is high and the consequences of infection severe. But what about a disease such as Lyme disease, where the risk of infection is relatively small, and the consequences not so deadly?

Lyme disease is transmitted by deer ticks, and is not transmitted person-to-person. The risk of infection is limited to areas where people share territory with deer, including swathes of central Europe and a growing envelope of rural and suburban North America. The disease is nasty but does not normally kill, and it can usually be cured by antibiotics. Confidence in the first Lyme disease vaccine stumbled

after 1999, when it became available in the United States. A campaign claiming that the vaccine caused side effects, including autoimmunity (see News Feature, page 524), caused sales to plummet, and the manufacturer GlaxoSmithKline withdrew the vaccine in 2002.

So what does a vaccine maker have to gain from trying again? Baxter Vaccines, based in Vienna, Austria, must be asking itself this question. It has invested sizeable resources in developing a new vaccine, and is considering whether to put it into clinical trials. Baxter is not the only company seeking a Lyme vaccine, so manufacturers are clearly convinced that a potential market exists.

One reason for that is the growing extent of the disease, especially in the sprawling suburbs of the United States, where the number of cases in areas with systematic surveillance has doubled since the early 1990s. Good statistics are not available for Europe, but one study in eastern Germany showed the incidence up by one-third between 2002 and 2003. In Austria, the disease is endemic, with 130 new cases each year per 100,000 people.

Baxter will be hoping that increasing risk to the public will reduce the aversion to a vaccine. US physicians note that those keenest to be vaccinated tend to have first-hand experience of the disease and its unpleasant treatment, which involves weeks of injections with powerful antibiotics. Baxter's candidate vaccine has been engineered to remove the part of a protein that the opponents to the vaccine held responsible for causing problems, even though the US Food and Drug Administration found no evidence for such harm.

It may go against the scientific grain for marketing considerations to play such a part in steering vaccine development. But in the real world, this may be unavoidable. Lyme disease is a serious illness and those who live in areas where it is spreading deserve a vaccine. ■

"Those who live in areas where Lyme disease is spreading deserve a vaccine."

Recycling the past

The reprocessing of nuclear fuel is an idea that should be laid to rest.

Plans to revive nuclear power are stirring on both sides of the Atlantic. In Britain, Tony Blair's government has been making upbeat noises about the need to replace existing nuclear power plants to fend off both national dependence on foreign sources of energy and global warming.

In the United States, however, President George Bush is said to be contemplating a step that will revive public concern about the link between nuclear energy and nuclear weapons — and could

ultimately set back any prospect of reviving the former.

When it is released next week, Bush's 2007 budget proposal is expected to include a provision that would start to revive nuclear-fuel reprocessing. That would end a three-decade-old strategy in the United States that has sought to sever the connection between nuclear power and nuclear weapons.

Nuclear-fuel reprocessing aims to reduce the volume of spent nuclear fuel that has to be disposed of safely by recycling it for use in new types of nuclear reactor. But the recycling involves separating components that can readily be used to build nuclear weapons.

Of the countries with significant nuclear power capacity, the United States and Germany abandoned reprocessing early on, and Britain, having ditched the fast-reactor design that would burn the recycled fuel, looks set to follow suit. Japan is trying to build a



reprocessing plant, but only France has stuck resolutely with fuel recycling. An official study commissioned by the French prime minister found recycling to be costly, however, and France has not yet managed to 'close' its fuel cycle by finding a place to put its waste.

The United States had (and has) ultimate responsibility for the nuclear-fuel cycle at plants that have been built by US contractors around the world. It abandoned reprocessing in a bid not just to lead by example, but to prevent a situation whereby countries that operate US reactor technology might obtain access to plutonium production lines.

The decision to abandon recycling sought to put the nuclear weapons genie back in the bottle in arguments over nuclear energy, in the United States at least. Bush's plan would release it again — and galvanize US opposition to nuclear power. Its adoption by Congress would effectively concede that US plans for the safe long-term disposal of nuclear waste at Yucca Mountain, Nevada, are not going to solve the waste problem.

The plan to revive the nuclear-fuel cycle comes at a peculiar time. The Pittsburgh-based company Westinghouse, which constructed most of these US-built plants, is being purchased by Toshiba for \$5 billion. This suggests that, in the eyes of some seasoned Japanese business executives at least, general global prospects for nuclear power are improving.

The case for a nuclear power revival has been well rehearsed. The global panic induced by the 1979 performances of Jane Fonda and Michael Douglas in *The China Syndrome* — and inflamed by the real-life version released at Three Mile Island in Pennsylvania 11 days later — is beginning to die down. European memories of

the 1986 Chernobyl accident are also fading.

Perhaps more to the point, the case now rests not on the specious grounds that nuclear energy will be immensely cheap, but on the rather more solid supposition that it is less bad than the alternatives. With coal causing global warming, oil and gas equated with dangerous energy dependency on outside suppliers, and renewable sources unable to produce the gigawattage that we apparently require, nuclear power is firmly back in the picture.

Yet the waste issue will need to be addressed before any ground is broken for a new nuclear power station in either Britain or America. Britain abandoned plans to build an underground waste repository in the north of England in 1997, and a report due this summer from a consultative panel, the Committee on Radioactive Waste Management, is only the first step in the search for a new approach. In the United States, the outlook for the Yucca Mountain project is uncertain, and the proposed repository there is, in any case, too small to meet forecast needs.

It may be that the Bush proposal reflects the administration's frustration over continued opposition to the Yucca Mountain repository. But, in the end, the only environmentally or financially viable path to nuclear power generation involves wrestling with the murky details of long-term waste disposal. Fuel recycling may look exciting on paper; in practice, it is part of the problem, not the solution. ■

"The case for nuclear energy now rests on the supposition that it is less bad than the alternatives."

Malaria quagmire

Progress in addressing Africa's largest health problem remains painfully slow.

Tackling malaria in Africa should not be beyond our means. For a few billion dollars a year, it ought to be possible to save the lives of millions of people and lift some of the world's poorest areas out of poverty.

All that is required, say specialists in the field, is access to tried and trusted remedies for those in Africa who need them most. Although vaccines and other new treatments would be helpful in the long term, some basic means of combating malaria, such as bednets impregnated with insecticide, are available now. Why aren't people getting them?

The economist Jeffrey Sachs is asking the question in his latest role as director of the UN Millennium Project, an ambitious initiative to reduce global poverty by 2015. At a meeting in Stockholm earlier this week, he gathered public-health officials, drug company executives, African politicians and others to try to find the answer.

Given the turf wars and squabbling that afflict this sphere, even the title of the conference — "A Malaria and Neglected Tropical Diseases Quick-Impact Initiative Meeting" — was likely to be greeted with scepticism, as the recent battle against malaria and other tropical

diseases has been characterized by neither speed nor impact.

The World Health Organization's Roll Back Malaria Initiative, which in 1998 pledged to halve the malaria burden by 2010, has struggled to establish itself and looks set to fall far short of its main goals.

Even so, the past decade has seen considerable progress in global public health. Malaria, tuberculosis and AIDS are high on the public agenda, and funding for control measures and for research has grown rapidly. On the ground in Africa there has been considerable, if uneven, progress.

The Stockholm meeting nevertheless provided ample evidence of the huge obstacles that remain in the way of implementing control measures. It was told, for example, that manufacturers can already produce 75 million bednets a year. But instead of agencies simply buying them and shipping them on, they have to pass through a tortuous circuit of tenders and approvals.

Charity Ngilu, Kenya's health minister, also pointed out that most African countries have next to no health infrastructure for the efficient distribution of drugs and bednets.

None of these problems will be addressed overnight, as the interminable nature of some of the discussion in Stockholm demonstrated. The world's attention must remain firmly focused on these diseases, however, until donor nations, African governments and international organizations find solutions, and achieve universal access to these basic control measures. ■

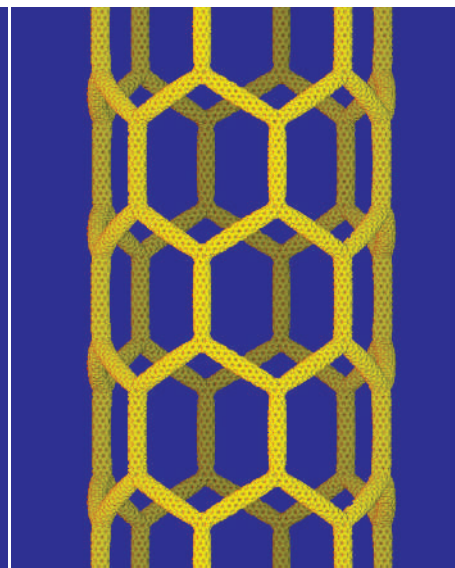
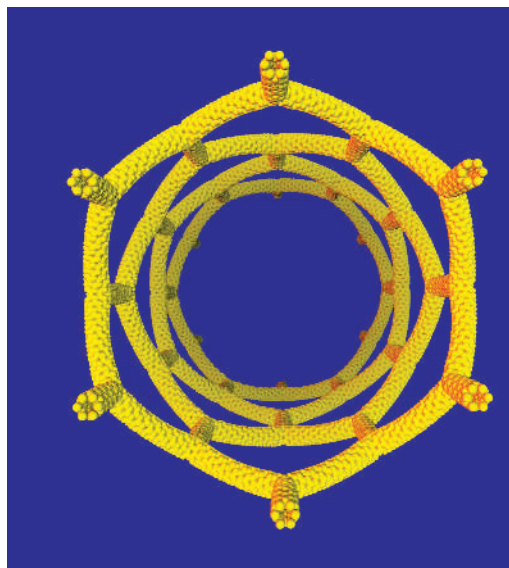
RESEARCH HIGHLIGHTS

Supertubes

Nanotechnology **17**, 617–621 (2006)

First there was just graphite and diamond. Then came fullerenes and nanotubes. Now theoretical physicists in Brazil introduce another form for carbon — the 'supernanotube'.

An ordinary nanotube is made from carbon atoms arranged in a hexagonal grid, in which each atom bonds to three others. The supernanotube proposed by Vitor Coluci of the University of Campinas and his co-workers has a similar structure (pictured), but with nanotubes in place of the bonds, resulting in Y-shaped junctions where the atoms would be. Using supertubes in place of nanotubes would create a supersupertube, and so on, to form ever-larger fractal structures.



V.R. COLUCI

CHEMISTRY

Order of the rings

J. Am. Chem. Soc. doi:10.1021/ja057908f (2006)

One of the great challenges in organic chemistry is developing novel ways to synthesize complex structures. The best reactions work for a range of chemical substrates, and give high yields.

By these criteria, a reaction, discovered by Derek Tan of the Memorial Sloan-Kettering Cancer Center in New York and his co-workers is near ideal. They report a new way to make spiroketals — unusual structures with two oxygen-containing rings that are connected by a single carbon atom.

The team obtained a number of spiroketals by reacting simple molecules with titanium salts. The oxygen atoms in the substrate seem to interact with the titanium in a way that helps to control which isomer is formed.

NEUROBIOLOGY

Some nerve

J. Cell Biol. doi:10.1083/jcb.200509174 (2006)

An evolutionary shift that occurred some 400 million years ago gave the myelin coats around the nerve cells of the central nervous system protective powers, according to a study led by Bruce Trapp at the Cleveland Clinic Foundation in Ohio.

The shift saw protein zero (P_0), which persists in the myelin of the peripheral nervous system, replaced in the central nervous system by the more complex proteolipid protein (PLP). Trapp's team 'reversed' this evolutionary step in mice, so that P_0 was expressed instead of PLP.

The swap halved the lifespan of the mice and degraded their motor skills. It also seemed to accelerate the normal age-related degeneration of myelinated nerve fibres.

ZOOLOGY

Record-breaking fish

Proc. R. Soc. Lond. B doi:10.1098/rspb.2005.3419 (2006)

Zoologists have unveiled the smallest free-living vertebrate ever found. Mature females of the fish *Paedocypris progenetica* (pictured below), which lives in highly acidic blackwater peat swamps in southeast Asia, average just 7.9 millimetres in length. Described by Ralf Britz, of London's Natural History Museum, and his colleagues, the miniature species has a larva-like appearance, with a skull that does not form properly, leaving it with a hole in the top. But the specimens' gonadal development shows them to be mature adults — and males possess a unique specialized structure near their genitals thought to be used for clasping the female during mating.

IMMUNOLOGY

Cytokine sound-off

Nature Immunol. doi:10.1038/nri1304 (2006)
Some immune cells shoot to kill, but helper

T cells sound the bugle to call in the cavalry. Mark Davis and his colleagues at Stanford University School of Medicine, California, show that they use a two-note alarm.

The cytokines secreted by helper T cells to activate immune cells seem to be distributed in two ways. One transport pathway targets antigen-presenting cells that interact with the helper T cell, the other distributes cytokines more broadly through the surroundings to recruit other immune cells. The pathways are associated with families of transport proteins known as Rabs and SNAREs.

The finding, says Davis, will shape ideas about the role that the delivery method plays in cytokine function.

PHOTONICS

Random sandwiches

Europhys. Lett. **73**, 225–231 (2006)

Distinguishing true randomness from something that just looks like it is hard — yet it can be vital.

To add to the raft of mathematical methods for assessing randomness, Alain Haché of the University of Moncton in New Brunswick, Canada, and his co-workers propose an ingenious approach based on physics.

They simulated light transmission through stacks of material in which the layers had



M. KOTTELAT/CORNOL, SWITZERLAND AND RAFFLES MUSEUM

different scattering properties. The transmission was strongly sensitive to disorder in the layers' properties. When the scattering properties encoded numbers, random sequences 100 numbers long could be distinguished from pseudo-random sequences with 70% fidelity.

A physical realization of the scheme should be possible using materials with tunable optical properties.

PHYSIOLOGY

Speed demons

J. Exp. Biol. **209**, 433–443 (2006)

Squid propel themselves through water at impressive speeds by relaxing and contracting their mantle — the muscular, tube-shaped structure near their head.

Baby oval squid (*Sepioteuthis lessoniana*) are also fast. But to swim quickly, these small creatures must contract their mantles at a much higher rate than adults of the same species. How do they do it?

Joe Thompson and William Kier of the University of North Carolina find that mantle muscle filaments in baby squid are, on average, 1.5 times shorter than those in juveniles and adults. Unlike vertebrates, which use biochemistry to fine-tune their muscle speed as they age, it seems that squid adjust the filament length.

EARTH SCIENCE

Meet the CHAMP

Geophys. J. Int. **164**, 319–330 (2006)

Earth's magnetic field is dominated by magnetism from the core, but certain minerals, such as magnetite in the planet's crust, or lithosphere, provide an additional contribution to the field. The CHAMP satellite has been measuring the lithospheric magnetic field since 2000. Its latest data highlight magnetic anomalies along subduction zones — where one tectonic plate sinks beneath another — that have never before been seen from orbit.

The image (above) shows the lithospheric magnetic field at an altitude of 50 kilometres, with subduction zones highlighted by thick green lines. Because the magnetic properties of the crust are temperature sensitive, the team of scientists led by Stefan Maus of the GeoResearch Centre Potsdam, Germany, hopes that precisely measuring and mapping the field will help to improve models of tectonic movement.

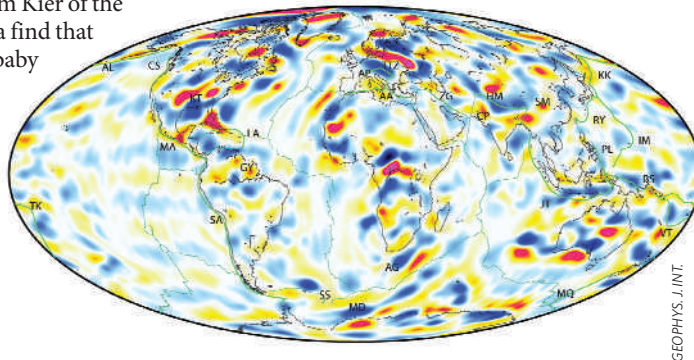
ASTROPHYSICS

Space intruders

Nature Phys. doi:10.1038/nphys214 (2006)

High-energy blasts from space known as short γ -ray bursts (GRBs) are thought to be produced when one neutron star merges with another, or with a black hole. But recent observations show short GRBs coming from above and below galaxies' main disks, where pairs of neutron stars, known as binaries, were expected to be less common.

One possibility is that GRBs originate in globular clusters, dense balls of ancient stars around galaxies' edges. Jonathan Grindlay of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues show that, in clusters, some binaries will form when one neutron star intrudes into another kind of binary, made of a neutron star and a low-mass star. Then neutron star mergers in clusters could be frequent enough to account for 10–30% of all short GRBs.



CELL BIOLOGY

Mitochondrial massacre

Proc. Natl Acad. Sci. USA **103**, 1382–1387 (2006)

Researchers in Japan have watched sperm mitochondrial DNA being destroyed in the egg immediately after fertilization.

Offspring inherit almost all of their mitochondria, which provide cells with energy, from their mother. This was long assumed to be because the handful of mitochondria delivered by the small sperm are diluted by the large egg's abundance of these organelles. But real-time observations made by Yoshiki Nishimura and his team back up reports that paternal mitochondrial DNA (mtDNA) is instead actively digested.

By fluorescently labelling the paternal mtDNA in fish, the researchers showed that it is broken down within hours of fertilization. This may prevent the embryo from inheriting paternal mtDNA that was damaged during sperm production.

JOURNAL CLUB

Carl Wunsch
Massachusetts Institute of
Technology, Cambridge, USA

An oceanographer charts the ebb and flow of opinion on ocean currents.

Many scientists believe that high-latitude cooling drives the ocean's currents as cold, dense water sinks then flows towards the Equator, creating a convective heat engine.

Since I was a student in the 1960s, when a colleague at MIT, Thomas Rossby, had seemingly shown that such a flow could be set up in the lab, this view has been entrenched.

But a flaw in the heat-engine model was pointed out as early as 1908 by the Swedish meteorologist Johan Sandström.

Heating a saucepan from below causes an instability: lower, warmer fluid rises and displaces the fluid above, leading to a vigorous convection current. In the ocean, heating and cooling both occur at the same level — the surface. Sandström argued that this situation should be stable.

In my student days, Sandström's argument was simply dismissed because he had considered an ideal, non-turbulent fluid.

Recently, however, some of us became interested again — in part because of public concern that the ocean circulation is 'shutting down', and more sensibly because of the need to understand the oceanic energy budget.

In one example, two experimentalists revisit the problem (W. Wang & R.-X. Huang *J. Fluid Mech.* **540**, 49–73; 2005) and find that cooling salty water at or below the level of heating always produces some motion. So Sandström wasn't strictly correct. But the observed flow is so weak that the efficiency with which the ocean converts heat to motion must be vanishingly small.

The circulation in the ocean — and in retrospect, in Rossby's original experiment — depends on details of how cold and warm waters mix. That means the winds and tides are the real drivers of the ocean currents. How long will it be until the literature catches up?

NEWS

Doubts over biochemist's data expose holes in Japanese fraud laws

TOKYO

An experiment published by biochemist Kazunari Taira is not reproducible, a University of Tokyo committee has concluded. The announcement, made on 27 January, vindicated many scientists who have challenged Taira's work.

But the ongoing investigation is raising more questions than it answers, as the committee seems unlikely to uncover what lies behind the invalid result, or address many of Taira's other papers, at least 12 of which have been questioned. Many in Japan are now calling for clearer guidelines to cover how universities should respond to cases of suspected fraud.

Taira is known for his work on RNA interference, in which small pieces of RNA intercept and regulate the process by which genes are turned into proteins. The molecules block specific genes, and are a powerful tool for studying gene function. Clinical use of the molecules to block the development of certain diseases is also getting under way.

Taira's work is not considered central to the promise of the technology. But so many scientists were trying — and failing — to repeat his work that the RNA Society of Japan became concerned. In April 2005, after several complaints from members, the society asked the University of Tokyo to examine 12 papers written by Taira and his team, including two *Nature* papers, one of which had already been retracted and the other corrected (H. Kawasaki and K. Taira *Nature* 423, 838–842; 2003 and H. Kawasaki and K. Taira *Nature* 431, 211–217; 2004).

The university set up a committee of internal and external experts to investigate. It asked Taira for raw data but he claimed that another author on the papers, Hiroaki Kawasaki, could not locate the proper notebooks. So the committee asked Taira to redo experiments described in four of the twelve papers by the end of December (see *Nature* 437, 461; 2005).

In a report submitted to the committee on 13 January, Taira said his team had reproduced one of these experiments, from a paper in *Nucleic Acids Research* (H. Kawasaki, E. Suyama, M. Iyo and K. Taira 31, 981–987; 2003). In that study, a human enzyme that can cut long RNA molecules into shorter pieces,

useful for RNA interference, is expressed and activated in the bacterium *Escherichia coli*. But the committee rejected the replicated result, partly because its own findings suggested that the experiment used different materials from those in the original work. An external company asked by Taira to repeat the test was also unable to activate the enzyme.

"Without concrete evidence, we could get sued."

Taira, who denies any wrongdoing on his part, saying that he is not capable of performing such "technical" experiments, blames Kawasaki for the discrepancies.

At a press conference on 27 January, a dejected Kawasaki spoke up only to say that the inconsistencies were just a matter of mislabelling. He has denied any wrongdoing.

Taira told *Nature* on Friday that he would not retract the paper, but will continue the tests. He is also trying to confirm experiments from a second paper; efforts to reproduce the third and fourth experiments haven't yet started.

Yoichiro Matsumoto, head of the investigation committee, says it will produce a final report in March. Another university committee may now meet to discuss possible disciplinary actions. But although Taira's credibility has suffered, a full investigation to determine whether fraud was involved is unlikely, because there is no clear legal authority for the

university to carry one out. Without concrete evidence, "we could get sued," says Matsumoto. Most countries either have a governmental oversight body to deal with possible fraud, or research institutions have guidelines in place to determine how to proceed.

Akira Yoshikawa, director of the science ministry's Science and Technology Policy Bureau, says that it is unlikely that Taira or Kawasaki will be investigated for criminal fraud — that is, for seeking funds based on research they knew to be fake — because the results so far have been "grey".

The situation is in stark contrast to the intense investigation that South Korea's Seoul National University (SNU) recently completed into the work of cloning researcher Woo Suk Hwang (see *Nature* 439, 122–123; 2006). There, the SNU conducted independent tests. The investigation team shut down Hwang's lab, examined notebooks and other materials, and conducted in-depth interviews, to reach a firm conclusion within a few weeks.

Jung-Hye Roe, head of research at SNU, says that although there was no clear legal framework in South Korea either, the university president gave the committee "full authority" to investigate — an unprecedented move motivated by a petition of university scientists.

The situation in Japan leaves researchers in the field uncertain about the merits of much of Taira's work. For example, Kunihiro Matsumoto, a biologist at Nagoya University, tried using Taira's technology to block gene activity, but failed to make it work. "I don't know if the problem is the technology or the genes we used," he says.

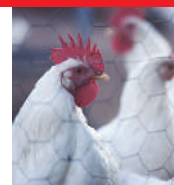
Many in Japan are now calling for clearer guidelines on dealing with fraud. In response to the Taira case, the government's Council for Science and Technology Policy says it will consider measures, although it's not clear yet what those would be. And the science ministry says a committee will report on the issue by the summer.

But neither is considering creating an independent body, such as the US Office of Research Integrity. "In some countries there is a government office to take care of such things," says Yoshikawa. But in Japan, we are already trying to cut back the number of government officials. The last thing we want to do is open up a new office." ■

Ichiko Fuyuno and David Cyranoski



RNA-interference researcher Kazunari Taira insists that he will be able to replicate his results.



QUICK VACCINE GETS OFF THE STARTING BLOCKS

Common cold helps grow swift protection against bird flu.

www.nature.com/news

SNAPSHOT Weather watch

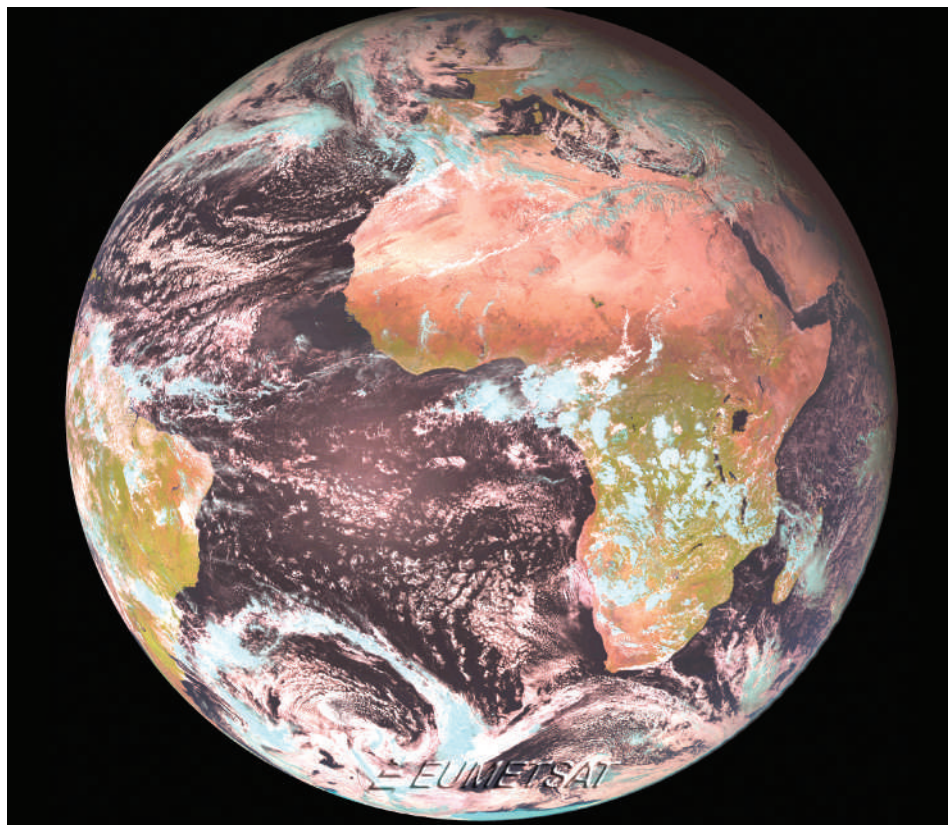
This false-colour infrared image of Earth comes from Europe's newest weather satellite, which was lofted into geostationary orbit on 21 December 2005. A swirl of icy cloud in the south Atlantic is shown in blue, and North Africa is shaded red.

The second Meteosat Second Generation satellite (MSG-2) will take images at infrared and visible wavelengths every 15 minutes, which will allow meteorologists to track changes in weather systems.

It will also measure the net balance between incoming radiation from the Sun and outgoing radiation from Earth much more accurately than satellites in lower orbits. Combining the results should reveal how cloud variation affects the heat that Earth loses to space, for example.

After an initial test phase, MSG-2 will enter seven years of operational service above the Gulf of Guinea, and will be renamed Meteosat-9.

Mark Peplow



Germany urges NASA to save airborne telescope

German space officials were pressing NASA last week to honour US commitment to the SOFIA (Stratospheric Observatory for Infrared Astronomy) project, following rumours that the agency may pull funding for the airborne observatory before its first flight.

SOFIA, which has been in the planning stage for more than two decades, will place a 2.5-metre telescope — similar in size to the Hubble Space Telescope — onboard a modified 747 aeroplane. This will fly above most of the water vapour in the atmosphere that blocks incoming infrared radiation. A successor to the Kuiper Airborne Observatory that flew until 1995, SOFIA bridges the gap between ground-based telescopes and infrared satellites such as the Spitzer Space Telescope. NASA will pay 80% of the bill, or about \$500 million, as well as most of the operational costs, with Germany supplying the telescope.

The first test flight of the long-delayed project was scheduled for this October, with science flights beginning in 2008. But at the American Astronomical Society meeting in Washington DC last month, project scientists heard rumours that NASA planned to eliminate funding for SOFIA in 2006 and 2007, with an eye to killing the project outright.

David Black, president of the Universities Space Research Association in Columbia, Maryland, which manages SOFIA for NASA, says the information came from many sources and is credible enough to be taken seriously. NASA spokeswoman Erica Hupp says the agency will not comment on SOFIA's status until the agency's 2007 budget request is unveiled on 6 February.

Alarmed by what they heard, German space officials last week

complained to their NASA counterparts. Sigmar Wittig, chairman of the board for the German Aerospace Center, DLR, pleaded SOFIA's case with NASA administrator Mike Griffin, as well as with deputy administrator Shana Dale, who was in Germany conferring

"The German card has been played — we've made very clear how important this is to us."

with science ministers and legislators. "The German card has been played," says Reinhard Genzel

of the Max Planck Institute for Extraterrestrial Physics in Garching, a co-investigator for one of SOFIA's planned science instruments. "We've made very clear how important this is to us." The German investment of nearly US\$100 million in SOFIA makes it the country's largest collaborative space project outside the European Space Agency.

The German pressure, along with queries to NASA from members of the US Congress, may already have

had an effect. "They're still working on it within NASA," says astrophysicist Eric Becklin of the University of California, Los Angeles, the SOFIA chief scientist. Although he was "prepared for the worst" earlier last month, Becklin now says of NASA's budget request: "We don't really know what's going to come."

Cancelling such a high-profile project so late in its development would be highly unusual. But NASA science managers are desperate to contain costs these days, and SOFIA's 20-year operational budget would make a tempting target. Having added substantially to the project's staff at NASA's Ames Research Center in California — partly to satisfy government-imposed safety requirements for flying the 747 — NASA is facing a lifetime operational budget for SOFIA that could top \$1 billion. "We are definitely looking at ways to cut the operations budget," says Becklin.

Tony Reichhardt

ON THE RECORD

“Public concern is probably the only thing capable of overcoming the special interests that have obfuscated the topic.”

Climate scientist James Hansen argues for open discussion of climate change, charging that NASA is trying to muzzle him for promoting cuts to greenhouse-gas emissions.

“No harm can possibly come out of eating good haggis.”

A Scottish butcher protests against official warnings that frequent haggis consumption could contribute to obesity in young children.

Sources: *New York Times*, *BBC*

SCORECARD

**Prison break**

Israel's Megiddo prison, together with its 1,200 inmates, looks set to be relocated after an ancient Christian church was unearthed in its grounds.

**Space suits**

Old space apparel is being given a new lease of life: as a satellite. NASA is to throw a space suit equipped with batteries and a radio transmitter out of the International Space Station. Ham radio operators are encouraged to track the 'SuitSat'.

**'Undercover' research**

A professor whose membership of the US National Socialist Movement got him fired from his university post claims he joined the Nazis as a research project to understand the mentality of white supremacists.

NUMBER CRUNCH

Chinese officials last week unveiled a plan for an aerial photographic survey to measure the Great Wall.

6,300 km is the length of the wall according to Ming dynasty historical documents.

7,000 km is its length according to local surveys last century.

US\$25 million is the price tag for the new survey.

Source: <http://english.epochtimes.com>

Experts plan to reclaim the web for pop science

Is it feasible to peer-review the Internet? A coalition of science agencies and Silicon Valley entrepreneurs is trying to do just that. They are launching what they claim will be an authoritative network of websites, where users can find trustworthy information on any subject. Top science organizations are signing up, but critics are sceptical about the project's rationale, and whether it can succeed.

The Digital Universe project is billed as a “network of web portals”, run by experts, on topics ranging from the Arctic and the oceans to the Solar System and the Universe. Users would navigate through the portals using a three-dimensional browser.

You could “fly over an accurate virtual Earth to explore the contours of the Grand Canyon, swim with the fish of the Great Barrier Reef and travel through the human body”, says an enthusiastic Robert Corell, chair of the steering committee for the Arctic Climate Impact Assessment and senior scientific adviser for the Digital Universe's Earth Portal.

The project also includes an encyclopaedia that will use similar technology to the popular online encyclopaedia Wikipedia, and Larry Sanger, a co-founder of Wikipedia, is helping to create it. But that's where the resemblance ends. All content in the Digital Universe will come from vetted experts, and articles will be reviewed by editors before going live. There will also be links to approved websites.

The driving force behind the project is ManyOne, a company headed by Joseph Firmage, who made a fortune in the 1990s from an Internet consulting company. He resigned in 1999 after the fallout from his book claiming that he had encountered extraterrestrials.

Firmage says he vehemently opposes the “anyone can edit” vision of Wikipedia. “Wikipedia is a very uninviting place for most intellectuals,” he adds. “I myself would not want to be writing articles that could be edited by somebody who does not necessarily have any expertise.” He hopes that peer-reviewed content will raise the standard of content on the web, which he describes as having “an intellectual deficiency of serious proportions”.

“The Digital Universe is an attempt to massively mobilize the scientific community,” adds Cutler Cleveland, director of the Center for Energy and Environmental Studies at Boston University and editor of the Digital Universe's Earth Portal and other portals on Earth's environment. “The information you see here you will know is trustworthy in a fundamental way.”

Many are enthusiastic. The US National Council for Science and the Environment, the University of California, Berkeley, Massachusetts Institute of Technology, the University of Oxford and the World Resources Institute have all signed up to help. The Digital Universe's Earth Portal, which is to be released next

“They're not in it for the money; actually, they're trying to save the world.”



The Digital Universe: science you can trust on the Internet.

month, has contributors that include Gene Likens, the discoverer of acid rain, Thomas Kunz, a top bat expert, and Robert Costanza, founder of the field of ecological economics. Its international advisory board includes Rita Colwell, former director of the US National Science Foundation.

Critics interviewed by *Nature* were unwilling to speak on the record. But some believe that the project is over-complicated, and that much of its underlying technology — which still requires significant development — runs against the trend to distribute information in lightweight formats that can be accessed by cell phones or PDAs such as the BlackBerry. “If you have to rely on a high-bandwidth always-on network environment, on devices with a lot of storage, you are pretty much going in the wrong direction,” says one critic, an expert in Internet information systems. He is also unimpressed by the Digital Universe’s concept

of peer-reviewing material. “There’s more than enough content on the web, even substantive content,” he says. “I’m not sure that generating new content is really a breakthrough.”

There are also questions over the business model, in which revenue would largely come from selling high-speed Internet access, with half the profits fed back into the work. “It’s an odd choice; that’s a dying business,” comments one observer familiar with the project, pointing out that in the future consumers will be unlikely to notice where their Internet access comes from. But he says he can’t help being inspired by the idea. “They’re trying to package science in a way that has some of the glitz and entertainment appeal of television, but that is also complete and correct,” he says. “They’re not in it for the money; actually, they’re trying to save the world.” ■

Declan Butler

♦ www.digitaluniverse.net

Senators seek cash to save US science

WASHINGTON DC

The United States is losing its scientific edge and needs billions of extra dollars to rekindle innovation, according to a bipartisan group of US senators.

After years of growth in areas such as biomedical research, funding at US science agencies is now mostly flat or decreasing, and critics charge that the nation’s competitiveness will soon suffer. The senators’ solution consists of three new bills that would dramatically increase the number of science teachers nationwide, boost funding for research, and increase tax breaks for industrial research and development.

“This is a basic problem that America faces and that everybody in the US Senate ought to be totally committed to solving,” said Senator Pete Domenici (Republican, New Mexico), unveiling the bills on 25 January. The other sponsors of the legislation are Democrats Jeff Bingaman (New Mexico) and Barbara Mikulski

(Maryland), and Republican Lamar Alexander (Tennessee).

The legislation, collectively dubbed the Protecting America’s Competitive Edge (PACE) Act, recommends spending US\$9.5 billion in the first year alone. That money would go towards hiring some 10,000 science and maths teachers, creating a crash programme for advanced energy research, and boosting funding by 10% at agencies such as the National Science Foundation, the Department of Energy and the National Oceanic and Atmospheric Administration.

Over the long term, the PACE Act recommends doubling the budget of these and other research offices, as well as providing research tax credits for industry and incentives for students who take a bachelor’s degree in science, maths or engineering. The legislation is an attempt to implement the recommendations made in a report by the National Academies, which was released last October and warned that other countries were

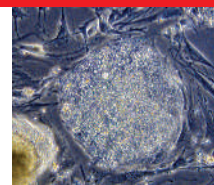
threatening US scientific dominance (see *Nature* 437, 1208; 2005).

Not surprisingly, science advocates were almost universally enthusiastic about the proposed legislation. “We’re excited about it,” says Gerald Wheeler, executive director of the National Science Teachers Association in Arlington, Virginia. “We’re creating the engine that’s going to get us out of this crisis.”

But others are more cautious about the bills’ prospects. Sam Rankin, president of the Coalition for National Science Funding in Washington DC, notes that the PACE Act would only recommend spending levels. Money for the initiative would have to be coaxed from congressional appropriators, who must also fund the US presence in Iraq and other government programmes. “Time will tell whether the rhetoric is backed up by the funding,” Rankin says. ■

Geoff Brumfiel

Additional reporting by Jacqueline Ruttimann



STEM CELLS IN FOCUS
Find a collection of all our news online.
www.nature.com/news/infocus/stemcells.html

Stem-cell tagging shows flaws

Biologists have overlooked problems with a decades-old technique for marking cells, warn US researchers. They say their findings call into question some studies of adult stem cells. The dispute is part of an ongoing and heated debate within the stem-cell community about whether adult stem cells extracted from one tissue, such as bone marrow, can spawn the cell types of a different tissue, and so be used to repair damage caused by disease.

Cell biologists have long relied on a technique called thymidine labelling, whereby a group of chemicals such as bromodeoxyuridine (BrdU) are incorporated into the DNA of dividing cells.

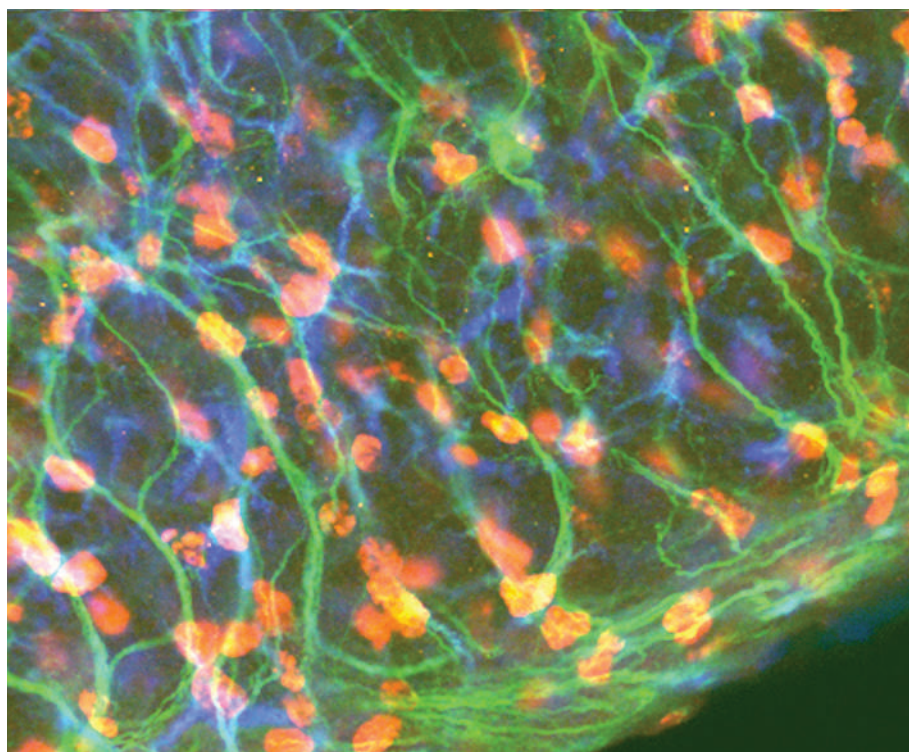
In some studies, BrdU is used to mark cells splitting at a particular time, allowing researchers to later examine where those cells migrated and what new cell types they generated. Often, stem cells are extracted from one tissue, marked with BrdU and then transplanted into a second tissue, such as the brain.

Now a team led by Catherine Verfaillie at the University of Minnesota, Minneapolis, is raising a red flag over the method. The group found evidence suggesting that many thymidine-labelled stem cells die after being transplanted into the brains of mice, and then release the chemical, which is taken up by neighbouring cells that are also dividing. This means that cells normally resident in the tissue could become marked and mistaken for transplanted cells.

The results were first published online in December 2005 (T. C. Burns *et al. Stem Cells* doi:10.1634/stemcells.2005-0463; 2005). "Hopefully, this will serve as a long-overdue wake-up call that will illustrate how misleading thymidine labelling can be," says Verfaillie.

Certain investigators who have used the method for many years say they already run rigorous control experiments that should avoid potential problems. One method is to transplant dead cells labelled with BrdU and compare the results to those obtained with live cells. Many researchers also use additional techniques to label their transplanted cells, such as green fluorescent protein (GFP), which is thought to be a reliable cell marker.

Even so, some experts say thymidine labelling is so commonplace that biologists may be blasé about potential pitfalls and skimp on the controls. "It's important to have a caution about these issues," says neurobiologist Fred Gage at the Salk Institute for Biological Studies in San Diego, California.



T. BURNS

Chemicals used to mark stem cells to track their behaviour may be taken up by cells not originally tagged.

Verfaillie says that the problems with BrdU could potentially affect any type of cell. But her study has focused on adult bone-marrow stem cells. Several researchers have reported that certain stem cells extracted from adult bone marrow and transplanted into the brains of mice can transform into new neurons or their supporting glial cells — fuelling the idea that these cells could repair brain disease or damage. Verfaillie says some of these results might be experimental artefacts, and that researchers should double-check their results: "In our hands it's all false positives," she says.

Cause to question

A study by Donald Phinney at Tulane University, New Orleans, on the properties of bone-marrow stem cells (G. C. Kopen, D. J. Prockop and D. G. Phinney *Proc. Natl Acad. Sci.* **96**, 10711–10716; 1999), is among those questioned in Verfaillie's paper. Phinney counters that since Verfaillie's study, many groups, using different techniques, have independently reported that transplanted bone-marrow stem cells acquire properties of neurons and glial cells, and questions whether the phenomenon Verfaillie's team saw could be unique to their

experiments. "I think it'll be controversial and raise the ire of a lot of people," he says.

There are additional, overlooked hazards of using BrdU, warns neuroscientist Pasko Rakic of Yale University in New Haven, Connecticut. One study from his lab showed that an injury can trigger neurons in the brain to start synthesizing new DNA, and that these cells can also take up the marker even though they are not actually dividing (C.-Y. Kuan *et al. J. Neurosci.* **24**, 10763–10772; 2004). A second study from another lab also suggested that BrdU could boost the ability of bone-marrow stem cells to generate new cell types. This implies that some of the properties previously attributed to the cells may actually have been triggered by the chemical altering gene activity (T. Y. Qu *et al. Rest. Neurol. Neurosci.* **22**, 459–468; 2004).

Such findings highlight the need for a broad rethink about how best to label transplanted cells, says Kenneth Chien, who studies heart stem cells at the University of California, San Diego. Biologists need "a more rigorous tool kit", he says — a repertoire of cellular markers that can unequivocally distinguish which cells have been transplanted, whether they are alive and precisely what they become.

Helen Pearson

"This will raise the ire of a lot of people."

SPECIAL REPORT

Should journals police scientific fraud?

Editors don't expect peer review to catch deliberate fakers. But recent scandals mean that journals are looking at other ways to detect fabricated papers. **Emma Marris** investigates.

The scientific community is sunk in one of its periodic bouts of angst over research fraud. After the Korean researcher Woo Suk Hwang's cloning work turned out to be spectacularly false, recent weeks have brought revelations that range from spurious Norwegians in a cancer study to doubts over RNA work in Japan (see page 514).

With high-profile cases putting the reputations of journals — and science as a whole — on the line, is there more that editors can or should do to prevent such embarrassments?

Whenever a published paper is exposed as a fake, editors are wont to repeat that peer review is not capable of catching fraudulent science. After a Norwegian study published in *The Lancet* was found to be based on imaginary patients (see *Nature* **439**, 248–249; 2006), the journal's editor, Richard Horton, pointed out: "Short of me flying to Oslo and checking out every entry on the computer, there is really no way for me to detect the fraud."

And other groups seem to agree that the primary responsibility for determining whether a paper ought to be shelved in fiction or non-fiction should not rest with journals. This opinion is based partly on the patchy staffing and funding of most journals, which are volunteer-run society publications. "Journals don't have the resources or the expertise," says Mary Scheetz, director of extramural research at the Maryland-based Office of Research Integrity, which investigates ethical violations in work funded by the US National Institutes of Health.

In addition, journals have very little disciplinary power — the worst they can do is refuse to publish work, or publish a retraction. The responsibility for investigating allegations lies mainly with institutions and funding agencies that pay for the work, points out James Kroll, who examines misconduct at the US National Science Foundation.

But in the past few years, journal editors have been taking a more proactive approach to dealing with fraud, and exploring what they can do with the resources they have.

Examining every paper submitted for fabrication would be pretty much impossible. "It

would be an astronomically expensive and difficult thing," says Drummond Rennie, deputy editor of the *Journal of the American Medical Association*. "It would take months; we are talking hundreds of thousands, and sometimes millions of dollars."

Stephen Evans, a statistician at the London School of Hygiene and Tropical Medicine, occasionally analyses papers in which the raw data are suspect. Tricks include looking for 'digit preference', the tendency of humans to round towards 0s and 5s, or the amount of variance in the data. "It is very difficult to invent data that has the right variability," says Evans. But he agrees that the time and expense make checking every study "totally impractical".

Instead, journals are investigating the potential of automated computer searches on submitted data, which could be incorporated into the review process with minimal time and effort. One idea catching hold is the introduction of screens to catch unacceptable image manipulation (see 'Forensic software traces



Paper trail: statistical tests may detect fake data, but checking every submission is impractical.

tweaks to images'). Editors are also exploring text-comparison software to help pick up plagiarism (see *Nature* **435**, 258–259; 2005).

"As information technology becomes more sophisticated, I think you are going to see more journals adding new tools to their

Forensic software traces tweaks to images

The editors of scientific journals could catch fraudulent images by using computer tools similar to those being developed for law enforcement and photojournalism, say computer scientists.

Recent fraud at the lab of stem-cell researcher Woo Suk Hwang (*Nature* **439**, 122–123; 2006) highlights the ease with which scientists can cook up fake images. In the South Korean team's 2005 *Science* paper, for example, two photographs in the same figure were found to be partial duplications.

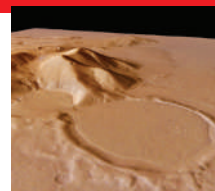
The Hwang furore is encouraging journal editors to seek ways of detecting

suspicious images before they are ever published. Such screening could identify outright fraud, as well as a far more pervasive practice in which biologists use programs such as Photoshop to tweak and smarten scientific images before publication (see *Nature* **434**, 952–953; 2005).

At least one group of journals, published by the Rockefeller University Press (RUP), already carries out such image forensics by eye. Mike Rossner, managing editor of the *Journal of Cell Biology* in New York City, has trained a production editor to enlarge and scrutinize images for obvious but

illegitimate changes, such as bands erased from a gel or cells slipped into a microscope image. This production editor checks manuscripts accepted for publication at the *Journal of Cell Biology* and two other RUP journals.

But this is time consuming, and computer scientists say that computer algorithms could automatically scan digital images and ferret out signs of manipulation. The development of such systems, prompted by the explosive spread of digital cameras and imaging programs, is also of interest to lawyers and police, who want to check for tampered crime-scene images, as well as to news organizations keen

**MARS ATTACK!**

Red planet under fire in proposed mission.

www.nature.com/news

ESA



screening processes," says Scheetz.

Another shift is that journal editors are now more likely to challenge papers that they think are suspicious, rather than quietly rejecting them. According to Harvey Marcovitch, head of the London-based Committee on Publication Ethics (COPE), the old way was to "find some excuse not to publish it". But the 200 or so journals that have signed up to COPE's code

of conduct are now committed to going further. "Even if you wouldn't accept a paper if it was completely clean, you have an absolute duty to inquire," says Marcovitch. "If you are not satisfied with the authors' responses, you have an absolute duty to go to the institution and ask them to investigate it."

Most would agree with that sentiment, but there are practical problems. In Britain, for example, editors are particularly loath to accuse anyone of fraud because the country's tough libel laws mean that they risk being taken to court. "Often what you do is ask for more and more clarification in the hope that something will turn up," says Marcovitch. A common step is to request the raw data, although this can be a headache for journals. "What do you do if you are suspicious about a paper, you ask to see the data and you get 25 cardboard boxes, 4 CDs and would have to hire a biostatistician for three months?" asks Marcovitch.

One simple step — which Scheetz says is taken surprisingly seldom — is to include misconduct policies in a journal's instructions to authors. "If something is suspicious, it puts the journal in a stronger position. Editors can request the raw data if it says they can in the instructions to authors," she explains.

Journals are also starting to request that researchers carry out their own checks before even submitting a paper. *Nature* now advises authors to include independent verification for certain cloning papers, for example. And

the *Journal of the American Medical Association* requires that industry-funded trials go through independent data analysis.

Journals are increasingly developing policies together, through programmes such as COPE, or using common policies drafted by groups such as the World Association of Medical Editors, based in Chicago, Illinois. The US Council of Science Editors, in Reston, Virginia, is currently working on a report on publication ethics. And *Nature*, *Science* and *Cell* are thinking about working up policies between them, according to Linda Miller, US executive editor of *Nature*.

Ultimately, if a journal does uncover evidence of fraud, it has to rely on the researchers' institution or funding agency to investigate fully. But this depends on such bodies having the will and authority to do so. When the *British Medical Journal* tried to get someone to investigate the work of cardiologist Ram Singh of Haldberg Hospital and Research Institute in Moradabad, India, for example, no institution or scientific body could be persuaded to make a judgment on the case. Singh went on to publish similar work in *The Lancet*. In the end, both journals published expressions of concern, but did not feel able to retract the papers. And in an ongoing case involving RNA researcher Kazunari Taira, the University of Tokyo seems unlikely to get to the bottom of whether suspicious data were faked, because it does not have the authority to make a full inquiry.

Emma Marris

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Artifice intelligence: programs can now detect manipulation in pictures.

to detect faked photos.

Partly funded by the FBI, Hany Farid at Dartmouth College in Hanover, New Hampshire, and his colleagues have designed a suite of ten mathematical techniques to

scan images for the hallmarks of manipulation. For example, one algorithm searches for small clusters of identical pixels in an image, and so might reveal where an area of background has been

copied and pasted over a blemish.

A second algorithm can identify whether part of an image has been expanded, perhaps to splice two photos seamlessly together. To do this, a program such as Photoshop generates new pixels by averaging the characteristics of the neighbouring ones — leaving a giveaway signature in the image.

Farid is currently converting his algorithms into a user-friendly form that can be attached to ImageJ, a free image-processing program distributed by the US National Institutes of Health. He is also consulting Adobe, the company that makes Photoshop, about whether the algorithms could be packaged together as a plug-in for the program, for use by different industries. This could help journal editors and reviewers, but also lab heads wanting to

check the work of their team.

Rossner says he plans to trial Farid's system when it is complete. Editors at *Nature* are also consulting researchers about automatic tools for detecting image manipulation.

But there are drawbacks. Such a system could act only as a first line of policing to flag up suspicious images. And editors and technophiles alike agree that anyone determined to fabricate or alter images will be able to fool the forensics software — perhaps by using the very same detection algorithms to learn how. Eric Postma is a computer scientist who devises software to detect fine-art fraud at the University of Maastricht in the Netherlands. "It's always a race between two sides," he says.

Helen Pearson

BSIP/SPL

NIH shells out to help postdocs find career path

Elias Zerhouni, director of the US National Institutes of Health (NIH), unveiled an award programme on 27 January aimed at easing the transition from postdoctoral scientist to independent researcher.

Many biomedical graduates drop out early in their careers, put off by insufficient funding and the competition for tenure-track faculty positions. Furthermore, postdocs are often not allowed to apply for independent research grants — such as the NIH's R01 award — which could help them secure a permanent job.

The NIH plans to issue between 150 and 200 of the awards — each worth up to \$1 million — starting this autumn and for the next five years. The awards will allow young investigators to complete their postdoctoral research with the first two years of funding, then move to a permanent position with the final three.

► http://grants.nih.gov/grants/new_investigators/index.htm

Dust stops comet watchers getting the hole picture

How big a hole did NASA's Deep Impact probe make when it crashed into comet Tempel 1? That is what the California-based Planetary Society asked the public to predict before the mission.

But six months after impact, astronomers can only guess that the crater is somewhere between 100 and 250 metres across. The dust simply never cleared in time for them to see.

So the society has chosen three winners at random from the 1,865 contestants whose guess was in that size range. Each will receive a plaque of the same copper used to make the impactor that hit the comet.

Moon dust escapes auctioneer's hammer

Collectors of space memorabilia may have missed out on the chance to acquire some cut-price Moon dust, after a journalist pointed out that the sample on sale was probably US government property.

Bonhams auctioneers in London were due to take bids on 24 January for a vial containing 31.2 milligrams of dust from the Moon. Lot 195, which was marked as having come from the Apollo landing of July 1969, had an estimated value of £300–400 (US\$530–700). Around 200 milligrams of lunar soil, collected by the Soviet Luna 16

Serbia airs plans for birthday honour to Tesla

Belgrade Airport will soon carry the name of inventor and scientist Nikola Tesla. Within weeks, the Republic of Serbia is expected to approve a plan to honour the 150th anniversary of Tesla's birth by renaming the country's largest airport as JP Aerodrom Beograd Nikola Tesla.

The move celebrates the country's connection with the brilliant and eccentric inventor, who was born in Croatia in 1856 to a family of Serbian origin. Tesla's initial research was conducted in Europe, but his fame stems from research he did after arriving in the United States in 1884. The industrialist George Westinghouse bought Tesla's patents for motors and dynamos that work with alternating current, and he used the devices to gain advantage over his rival Thomas Edison.

Tesla also claimed to have been contacted by



extraterrestrial life and to have invented a death ray capable of destroying 10,000 aircraft at a distance of 400 kilometres.

David Bowie is slated to play Tesla in the forthcoming movie *The Prestige*.

mission in 1970, sold for \$442,000 at auction in 1993.

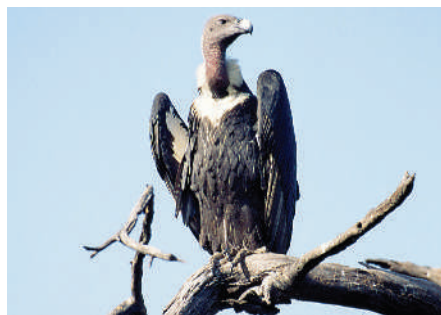
But any chance of a lucky bidder securing a bargain disappeared shortly before the sale. Bonhams withdrew the item after being contacted by a reporter from *The Economist* who pointed out that most lunar samples from NASA missions are considered US government property. Bonhams says it is investigating the matter with NASA and the US government.

Switching drugs may save vultures, study suggests

Conservationists are using a discovery about a drug used on livestock to urge India's government to address the falling numbers of Asian vultures (*Gyps*) in the subcontinent.

New research (G. Swan *et al.* *PLoS Biol.* 4, e66; 2006) shows that the anti-inflammatory drug meloxicam, which has the same activity in livestock as the widely prescribed drug diclofenac, does not cause toxic side effects in vultures. The birds can suffer kidney damage and gout when they eat the carcass of an animal previously treated with diclofenac.

Since diclofenac was introduced to treat livestock in the Indian subcontinent in 1994,



The bitter bit: many Asian vultures have died after eating animals previously treated with diclofenac.

the vulture population has plummeted by 95%. The Indian government pledged last March to ban the drug by September 2005, but the country is believed to have millions of stockpiled doses.

Government officials are meeting with conservationists in Delhi this week to discuss the vultures' plight.

Voluntary scheme aims to eliminate process chemical

The US Environmental Protection Agency has launched a voluntary programme to reduce the use and emission of an industrial chemical involved in making non-stick coatings, such as Teflon, and stain-resistant materials.

The chemical in question is perfluorooctanoic acid (PFOA), which does not readily degrade in the environment; its health effects are still being debated. The drive aims to cut both PFOA emissions and its appearance in products to 95% of 2000 levels by 2010, and to eliminate it by 2015. Teflon manufacturer DuPont, which has its headquarters in Wilmington, Delaware, has already signed up to the scheme.

PFOA is used as a processing aid during manufacture and is not part of the products. DuPont says that it will increase existing efforts to filter wayward PFOA out of air and water leaving its facilities, and also to recycle the compound.

Correction

The News story "Panel quits in row over sonar damage" (*Nature* 439, 376–377; 2006) incorrectly described a report released in January 2006 by the US National Marine Fisheries Service on the deaths of stranded whales as "final". It should have been described as "updated".



UPHILL STRUGGLE

The first vaccine against Lyme disease was withdrawn because patients distrusted it. Should market forces be allowed to shape the next one, asks **Alison Abbott**.

P. SCHERMEISTER/CORBIS

The idea that the customer is always right is gospel in most areas of business. Most, you might think, except the business of making drugs and vaccines. Not so. The cautionary tale of the Lyme-disease vaccine is a good example of how consumer power can override science.

Lyme disease is a debilitating infection spread by bites from ticks. Found mainly in the United States and Europe, its prevalence is increasing as civilization encroaches on forested areas — home to the deer on which the ticks feed. A vaccine is badly needed to contain the disease, but the one jab available was pulled from the market after complaints from a group of patients damaged its reputation. Now, a similar vaccine is set to emerge, but some say it has been tweaked in a way that confers no obvious scientific benefit.

Lyme disease made a meteoric entry into medical and public consciousness 30 years ago, when Allen Steere, a rheumatologist now at Massachusetts General Hospital in Boston, described an outbreak of a mysterious illness in Lyme, Connecticut. The disease he identified was caused by a spirochete, a spiral-shaped bacterium called *Borrelia burgdorferi*. It is rarely fatal, but acute infection results in a range of unpleasant and debilitating symptoms, including flu-like effects, a bullseye rash and neurological problems such as facial palsy.

After several weeks without treatment, some patients go on to develop joint pain or arthritis. If left untreated, it can last for years, but even then will usually respond to antibiotics. In a small subset of patients, however, the joint pain and swelling do not respond — the condition is known as treatment-resistant Lyme arthritis.

It was arguments about the long-term

persistence of Lyme disease that triggered the first 'Lyme scandal'. In the 1990s, Steere was hounded by patients who claimed to be suffering from a chronic form of Lyme disease that left them persistently exhausted. Steere insisted that they had been misdiagnosed. Although the patients had antibodies to Lyme disease, indicating that they had at one time been exposed to *Borrelia*, Steere maintained that their range of vague symptoms did not correlate with the true course of the disease. He says he wanted to save people from unnecessary antibiotic treatment, but instead found himself under protection of security guards, and dealing with hate mail and even death threats.

Ticked off

The dust had barely settled when another scandal broke, which eventually led Smith-Kline Beecham (now GlaxoSmithKline) to withdraw its vaccine LYMERix from the US market. Shortly after the jab was introduced, hundreds of vaccine recipients claimed that they had fallen victim to side effects — including autoimmunity, in which the immune system attacks the body. Their claims were based on a scientific hypothesis that is still unproven today: it predicts that people will raise antibodies to a stretch of a particular protein on the outer surface of *Borrelia*, called OspA, that could destroy normal human protein as well as

the bacterium. LYMERix happened to work by generating antibodies to this OspA protein — which is why patient advocacy groups latched on to the idea that the jab itself could cause autoimmune disease.

One law firm filed a class action in December 1999, and individual suits began to flood in. The US Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention published their own investigation into 900 or so adverse event reports and concluded that there was no evidence of a link between the autoimmunity complaints and the vaccine¹. But the bad publicity torpedoed sales. In February 2002, GlaxoSmithKline withdrew the vaccine from the US market and abandoned plans to launch a similar one in Europe. Enthusiasm for the jab seemed to have been killed off, and various attempts to generate vaccines against different *Borrelia* proteins fizzled to nothing.

But last year the company Baxter Vaccines in Vienna quietly announced that it was hoping to do clinical tests of a new vaccine candidate in Europe. To the astonishment of Lyme-disease researchers, the candidate turned out to be broadly based on the same *Borrelia* protein, OspA. The vaccine, which has been successfully tested in animals, differs from LYMERix in two ways. Europe is home to two more species of *Borrelia* than the United States, so Baxter's vaccine is designed to protect against infection from all three. And the short stretch of the OspA protein that had been associated with the hypothetical danger of autoimmunity has been spliced out.

The news only deepened the frustration of Markus Simon of the Max Planck Institute for Immunobiology in Freiburg, Germany. Along

"This just shows how irrational the world can be — there was no scientific justification for pulling the first Lyme vaccine."

— Markus Simon

First in line: those who work or play in the forest could benefit a lot from a vaccine for Lyme disease.

with his colleagues, Simon was the one who had developed the concept of an OspA-based vaccine in the early 1990s, which was then taken to the clinic by GlaxoSmithKline. "This just shows how irrational the world can be," says Simon. "There was no scientific justification for the first OspA vaccine being pulled."

On the trail

The idea that Lyme infection — and the Lyme vaccine — could cause the immune system to attack the body has its roots in the 1970s, when Steere described treatment-resistant Lyme arthritis. Nothing to do with the 'chronic Lyme disease' that angry patients later claimed had been triggered by Lyme infection, this syndrome, which can last for years, is characterized solely by inflamed joints. The arthritis seems not to be associated with continuing bacterial infection. Indeed, in the decades that followed, bacterial DNA has only very rarely been found in fluid extracted from affected joints, even using modern, sensitive methods of amplifying DNA traces². So scientists began to entertain the thought that this treatment-resistant Lyme arthritis could be an autoimmune response.

The idea was supported by the discovery in 1990 that the immune systems of most patients developing treatment-resistant Lyme arthritis shared some very specific, and genetically determined, characteristics³. This subset of people happens also to be particularly susceptible to rheumatoid arthritis, an autoimmune condition afflicting the joints.

If persistent Lyme arthritis were really an

"It is essential to err on the side of caution, and it is a simple matter to eliminate the problem."

— Brigitte Huber

autoimmune condition, what might be the mechanism? Scientists first looked to the idea of 'molecular mimicry', which was developed in the 1980s to explain how microbes in general might cause autoimmunity. According to this hypothesis, part of one of a microbe's proteins might be structurally similar to part of a normal protein in the patient — but different enough to be recognized as foreign by the patient's immune system. On exposure, antibodies to the foreign protein would be produced, which would then also attack the patient's normal protein.

To test whether the molecular-mimicry hypothesis might be applicable to treatment-resistant Lyme arthritis, scientists began to look for a host protein that the immune system might mistake for a *Borrelia* one. Looking for matches to OspA was an obvious starting point, and in 1998 Brigitte Huber, an immunologist at Tufts University in Boston, and a number of her colleagues including Steere, came up with a strong candidate. They found that a stretch of the *Borrelia* OspA protein shared a very similar amino-acid sequence with a human protein that helps move immune cells from blood vessels to inflamed tissue⁴. In the test-tube, this protein did the right thing — it prompted immune cells from treatment-resistant Lyme arthritis patients to trigger inflammation processes.

The implications of this for a vaccine whose

mechanism depended on OspA antibody production were clear. And patient advocacy groups, sensitized by the battle over 'chronic Lyme disease', were on the alert when LYMERix was approved for the US market by the FDA later in 1998.

But although molecular mimicry has a sound and respectable foundation as a hypothesis, no one has yet shown that it happens in real life. And as time went by, Lyme-disease researchers showed that many other proteins also activated OspA-specific immune cells in the test-tube⁵, and some of them did not even share similar sequences. "This all dealt a big blow to the molecular mimicry theory," says Thomas Kamradt, an immunologist at the University of Jena in Germany.

No verdict

Huber says that the jury is still out because there is not yet a good animal model in which to test the hypothesis. But most Lyme researchers are now convinced that molecular mimicry is an extremely unlikely explanation for any autoimmune response that might underlie treatment-resistant Lyme arthritis.

So, given that costly testing for safety and efficacy will have to start from scratch, does it make any sense to modify an earlier vaccine's structure in ways that may not, after all, be necessary? Simon thinks not. "We had years of clinical data with LYMERix — hundreds of thousands were vaccinated, and not one autoimmune response was ever confirmed," he storms. "And who knows if the new vaccine with this sequence spliced out will still be protective in humans?" Only clinical trials will tell. When contacted by *Nature*, Baxter Vaccines declined to comment.

Although no link has been found between LYMERix and an autoimmune response, some researchers are taking a precautionary stance. Steere, who has worked as a consultant for Baxter Vaccines, says: "There is no proof that autoimmunity ever developed in anyone, but it could be a very rare side effect." Splicing out the sequence in question "takes care of this theoretical concern", he adds. Huber agrees: "It is essential to err on the side of caution, and it is a simple matter to eliminate the potential problem."

Steere, whose experience with the angry patients encourages him to keep his head well below the parapet, strongly believes that a vaccine is necessary. And he argues that this need will become clearer to the public as experience of Lyme disease expands: "Most people who want a vaccine are those who have already had the disease."

Alison Abbott is *Nature's* senior European correspondent.

1. Food and Drug Administration www.fda.gov/ohrms/dockets/ac/01/briefing/3680b2_06.doc
2. Nocton, J. J. et al. *N. Engl. J. Med.* **330**, 229–234 (1994).
3. Steere, A. C., Dwyer, E. & Winchester, R. N. *Engl. J. Med.* **323**, 219–223 (1990).
4. Gross, D. M. et al. *Science* **281**, 703–706 (1998).
5. Maier, B. et al. *Eur. J. Immunol.* **30**, 448–457 (2000).



CDC/J. GATHANY



HIGH AND DRY

Two decades after plans were set in motion for the world's most powerful ground-based telescope, astronomers are bracing themselves for a downgrade to curb escalating costs. **Jeff Kanipe** reports.

The view from the sprawling Llano de Chajnantor plateau perched in the Andes Mountains of northern Chile is spectacular all year round. The plateau lies some 5,000 metres above sea level, with the skies above forming a deep blue backdrop to the tawny, windswept desert floor and the distant badlands of low hills and volcanoes.

Inhospitable to life, the plateau has become destination number one for radio astronomers, who probe the mysteries of the Universe by measuring radio waves emitted by celestial objects (see 'Transformational telescopes' on page 528). The high altitude and dry atmosphere are perfect for observing objects that emit submillimetre wavelengths; these wavelengths are in the spectral window that lies between the far infrared and high-frequency radio bands. Some of these wavelengths are absorbed by water vapour, ruling out observations at lower altitudes. Radio astronomers have been hoping for decades to be able to explore this window with the right telescope at the right location.

They will eventually get their wish when the international Atacama Large Millimeter Array (ALMA) is built on the Chajnantor plateau. By combining 64 radio antennas, each 12 metres in diameter, ALMA was originally projected to be 10 to 100 times more sensitive than current high-altitude telescopes. During a meeting with astronomers in Washington DC last

month, Adrian Russell, ALMA's US project director, described recent progress on the project, almost as if he could not believe it himself. "ALMA is actually happening," Russell announced. "It's real."

Attendees at the American Astronomical Society meeting had come to hear Russell and Fred Lo, director of the National Radio Astronomy Observatory (NRAO), which is responsible for US construction and operations, update them on the status and scientific prospects of ALMA. After their talks, Lo and Russell fielded technical questions from the audience, but one caught them flat-footed: how much was the project going to cost? Lo replied that because the National Science Foundation (NSF) was reviewing ALMA's US funding, it would be "damaging to the review process" to say more at this stage.

What went unsaid is the fact that ballooning costs for commodities such as petroleum and steel are forcing ALMA scientists to consider the unthinkable: whether to reduce the size and sensitivity of their dream telescope. Instead of the hoped-for 64 antennas,

they may have to do with 50 or even fewer.

The seeds for ALMA were sown in 1982, when an NSF committee proposed building an array of antennas up to two kilometres across, which would operate at millimetre wavelengths. A decade later, the National Research Council gave the project — then called the Millimeter Array — its blessing. In that incarnation, the array would have consisted of 40 8-metre radio dishes that could be switched from an array 70 metres across to one 3 kilometres in diameter.

Plans afoot

It is not unusual for radio arrays to have movable antennas. Changing the diameter of the array affects the fine detail seen by the telescope — what astronomers call angular resolution — and the whole movable array also works a bit like a zoom lens. For example, astronomers who want to map the large-scale structure of a galaxy use the smallest configuration; otherwise they would have a limited viewing angle (like looking at a freckle on someone's nose rather than on their whole face). If an interesting 'freckle' is detected then they can 'zoom' in and observe it in detail using the larger configuration.

Around the time the United States was planning the Millimeter Array project, European astronomers were discussing plans for an array

"It is critical that we do this instrument right, not halfway." —
Lee Mundy



The Llano de Chajnantor plateau in Chile, one of the highest, driest places on Earth, is a perfect viewing spot for radio astronomers.

in the Southern Hemisphere — the Large Southern Array — whose combined resolution would be equivalent to a dish 10 kilometres across. Despite the fact that the European array was not originally intended to work in the crucial submillimetre range, both sides began discussing a partnership in 1997. Two years later, an agreement was signed to merge the projects into ALMA. Today, Canada, Chile, Japan and Taiwan have all joined the collaboration, making it one of the first truly global projects in ground-based astronomy. And like most big projects it has a big price tag — about \$650 million — making it, so far, the most costly ground-based telescope ever.

ALMA's specialty will be observing cooler parts of the Universe the details of which are easiest to make out in the submillimetre spectrum. From interstellar dust to star-forming giant molecular clouds, many cold features in the cosmos have been glimpsed only by using high-altitude telescopes and special imaging techniques, or with long observation times. Scientists predict that ALMA will be able to image galaxies that formed up to 500 million years after the Big Bang — when the Universe was dark and star formation was just beginning. "These galaxies are truly spectacular, with star formation rates up to a thousand times that of the Milky Way today," says Jason Glenn, an astronomer at the University of Colorado and ALMA science advisory committee member. ALMA will give astronomers data on how fast the galaxies formed, and on the evolving structure of the Universe.

Most of ALMA's multi-million-dollar budget will be spent on antenna construction. The original plan called for enough antennas to form an array diameter that ranges from 150

metres to 14 kilometres across. In this plan, the smaller array size could be used to detect planet-forming disks around multiple stars in a galaxy, whereas the larger array would allow astronomers to scrutinize each disk more carefully, say, for the presence of Jupiter-size planets.

Price hike

ALMA's construction phase started in earnest in November 2003. But in 2004, the estimated costs started to balloon. Chinese demand for structural steel drove steel prices through the roof. Petroleum prices, too, more than doubled, inflating the cost of antenna dishes made from petroleum products. Even a boom in the Chilean economy translated to increased operational costs in that country.

The NSF, which is responsible for half of ALMA's construction costs, has to spread its funding across several high-priority projects, including the IceCube Neutrino Observatory at the South Pole and the EarthScope project in North America. Although ALMA's share jumped from \$29.8 million in 2003 to about \$50

million in 2004, it hasn't risen since. With commodity prices unlikely to decline anytime soon, this means that ALMA could easily spend its \$650 million before all the antennas are built. Facing the prospect of cost overruns, the NSF decided late last year to do a spending review.

The chances that ALMA will be downsized to 50 or even 40 antennas alarms some radio astronomers. Fewer antennas will mean reduced sensitivity and much longer exposure times to obtain the same resolution. Last year, the US National Academy of Sciences issued a report on the effect a smaller array would have on ALMA's three primary science goals.

The first of these, to detect Milky-Way-type galaxies that formed two to three billion years after the Big Bang, would help trace these galaxies' evolution — something about which astronomers know next to nothing. The second, to obtain high-contrast images of nearby protostars and planet-forming disks around young stars, would help astronomers study the formation of these cosmic structures. And the third, to produce precise images at very high angular resolution, would help reveal the motions of gas within objects such as giant molecular clouds.

The academy's report was discouraging. In their view, the first two requirements for these goals, involving sensitivity and high-contrast imaging, would not be met by either a 40- or 50-antenna array. With 40 antennas, the third requirement would also be at risk, even if extremely long exposure times were allowed. Reduced speed and image fidelity would be among the biggest challenges. For example, the time taken to observe a nearby stellar disk would increase from three days (with 64 antennas) to a week (with 40), and



Price cut: as costs rise, the number of antennas proposed for the new ALMA telescope is shrinking.

Transformational telescopes

Here are just a few of the radio telescopes responsible for making discoveries that transformed science. They were all created at a fraction of ALMA's cost.

Bruce Array Karl Jansky's primitive rotating radio telescope was the first to detect radio waves emitted by the Milky Way, back in 1933. Cost: unknown.

Reber's Antenna The world's first parabolic dish, 9 metres across, was built by Grote Reber with his own money. In 1938, he used it to confirm Jansky's discovery and to detect radiation from charged particles. Cost: \$4,000 (about \$45,000 today after inflation).

64-metre Parkes Radio Telescope, Australia In 1961, this telescope pinpointed the location of a radio source that was later confirmed to be the first known quasar. Cost: \$1.2 million (\$6.2 million today after inflation).

Bell Labs Holmdel Horn In 1963, astronomers Arno Penzias and Robert Wilson first thought pigeon droppings might be slightly raising their antenna's temperature. It turned out to be the cosmic microwave background radiation. Cost: \$25,000 for the computer; a Palm Pilot would suffice today.

305-metre Arecibo dish in Puerto Rico This impressive bowl-shaped antenna was used in 1965 to show that Mercury rotated every 59 days — not 88. It was also used to discover the first binary neutron star

and gravitational radiation in 1974. Cost: \$9.3 million; \$100 million to rebuild today.

University of Cambridge 1.8-hectare array Designed by Antony Hewish, this telescope was built using 190 kilometres of wire and cable, and used to discover the first pulsar in 1967. Cost: £15,000 (about \$303,000 today after inflation).

Arizona University's 36-Foot Telescope Now called the 12-Meter Telescope, it broke new ground in millimetre-wavelength molecular astronomy. Dozens of molecular species were first detected in interstellar gas by this instrument in the 1970s. Cost: \$1.5 million (\$7 million today after inflation).

Ohio State University's Radio Telescope Known fondly as 'the Big Ear', this instrument was used between 1965 and 1971 to map 20,000 radio sources, 60% of which had never been observed before. It was torn down in 1998 to make way for a golf course. Cost: \$250,000; about \$10 million to duplicate today.

Very Large Array The discoveries made using this telescope are legion. They include a mini-spiral of hot gas at the centre of the Milky Way (1983); traces of water ice on Mercury (1991); 'micro quasars' in the Milky Way (1994); and the detection of the first strong outburst from a magnetar, a supermagnetic neutron star (2005). Cost: \$78.6 million (\$133 million today after inflation).

image fidelity would be reduced threefold.

Despite the finding that it would fail to complete two of its three goals, the report concluded that a pruned-back ALMA could still address many unanswered questions in astronomy. Michael Turner, the NSF's assistant director for mathematical and physical sciences, is similarly hopeful: "The original ALMA plan has so much more capability over what already exists — by more than an order of magnitude, closer to two orders of magnitude — that even if you cut back by 40% what it can do, it's still transformational."

But if ALMA is to be truly transformational, "it is critical that we do this instrument right, not halfway," argues Lee Mundy, a member of ALMA's science advisory committee and a radio astronomer at the University of Maryland. According to Mundy, "chipping away at the number of antennas and collecting area saves dollars but it doesn't make sense for an instrument that can produce premier science for the next 30 years and more."

Lucy Ziurys, an astrochemist with the University of Arizona, says that a 50-antenna array is "okay, but disappointing." Although she believes that reducing the number of antennas is a better response to funding problems than reducing, say, antenna quality, she questions why certain costly decisions were made to begin with. "Funds could have been saved if only one antenna design had been chosen instead of two," she says. "That decision will, in all likelihood, increase the cost."

Ziurys is referring to two contracts signed last year. On the US side, the defense contractor General Dynamics, based in Falls Church, Virginia, agreed to supply 25 antennas in July. A similar contract was signed in December between the European Southern Observatory

and Alcatel Alenia Space, a European consortium of manufacturers. Ziurys argues that having two antenna designs means maintaining "two of everything" — two sets of spare components and electronics.

Unknown Universe

This coming March, the international ALMA board will meet in Kyoto, Japan, to finalize a delivery schedule for the rest of the project. That's when the NSF's decision to finance 50, or fewer, antennas will be made public. Last October, however, ALMA scientists reconfigured the antenna array at the Chajnantor site for 50 antennas. So the decision may have already been made. Ziurys thinks the decision to go with 50 was probably made as early as January last year.

Although their research may be adversely affected, most astronomers associated with ALMA seem resigned to the smaller array. Many remain hopeful that they can still

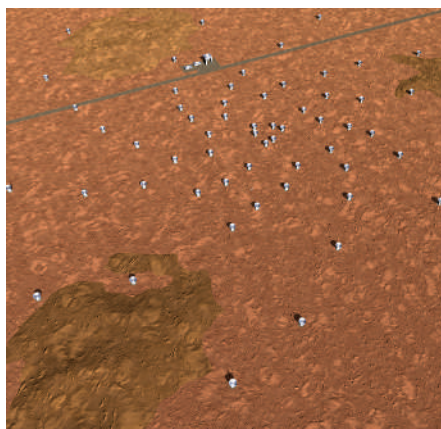
achieve worthwhile science with 50 antennas. "In general, science goals do not become impossible when the number of antennas is reduced slightly, they just take more observing time," claims Glenn. George Djorgovski, an astrophysicist at the California Institute of Technology, Pasadena, sums it up with a prediction: "What you have here is clearly a fiscal reality issue, and the politically considered response to it. We will see many more of those in the coming years."

Meanwhile, ALMA continues to move forward. Construction is under way at the site and the first of the antennas is scheduled to arrive in early 2007, with completion scheduled for 2012. For its part, the National Astronomical Observatory of Japan has agreed to build four 12-metre and twelve 7-metre antennas to create the Atacama Compact Array alongside ALMA. This will be used for imaging large-scale structures not well sampled by the main array.

Once the decision to go with 50 antennas has been made, could the project planners still build 64 at a later date? Not without an astronomical fairy godmother. "Buying more antennas would be hard," Turner explains. "You configure the array to work optimally based on how many antennas you have, and to add more would force you to reconfigure the entire array." That costs more money, which will be in short supply.

Will these few clouds spoil the view from Chajnantor? Paul Vanden Bout, former ALMA director, argues that new discoveries will erase any bleak memories. "Big projects are a long haul," he says. "But it's worth it. ALMA is going to be a great telescope and it's going to do exceptional science."

Jeff Kanipe is a writer based in Maryland.



The resolving power of the new telescope will be altered by changing the size of the antenna array.

TIME TO RAISE THE DEVIL

A horrible facial cancer is decimating the Tasmanian devil population. But researchers in Australia think they have found a way to save the species. **Carina Dennis** reports.

Checking her research traps at dawn, Billie Lazenby was saddened to find a female Tasmanian devil with a face marred by cancerous lumps. She had contracted devil facial-tumour disease, a deadly transmissible cancer that is threatening the survival of this feisty marsupial.

When reviewing her records later that night, Lazenby, who is part of a scientific team monitoring the devils, was astonished to find that the animal, nicknamed Half pea, had been recorded nearly a year previously as having the disease. This made her one of the longest survivors the team had ever encountered. The fact that Half pea had resisted the tumour for at least twice as long as most other devils meant she might hold a clue for scientists trying to help the species. "Amid this heart-breaking background of losing so many devils, suddenly this animal pops up and it's like, 'wow, maybe there is hope,'" says Lazenby.

Since this discovery in June 2005, the team has found two other partly resistant females, and researchers now hope to identify the genetic and immunological factors that give devils (*Sarcophilus harrisi*) a better chance of fighting the disease. It is a long-awaited piece of good news after a decade of helplessly watching the animals decline. Although long-term strategies for tackling the problem are some way off, other teams have discovered a way to limit the spread of the disease at least.

Deadly love bites

Lazenby, a scientific officer at the Tasmanian state government's Department of Primary Industries, Water and Environment (DPIWE), based in Hobart, is part of a team led by Clare Hawkins that traps and checks devils for signs of the cancer. A severely diseased devil is a grotesque sight: large tumours protrude from the face and neck, sometimes pushing out teeth and invading eye sockets. As the tumours interfere with feeding, the animals become emaciated and usually die within six months of showing lesions.

First documented by a wildlife photographer in 1996, the cancer has contributed to the deaths of up to 80% of infected devil populations in Tasmania¹, the only place in the world where the animal is found. Devils normally have a life expectancy of about five years but "now it's rare to see an animal older than



Simple measures could buy time for researchers working out how to beat a fatal disease in Tasmanian devils (below).



three years of age", says Lazenby.

So far, the evidence points to the disease being transmitted by infectious cancerous cells when animals bite each others' faces in fights or during their tempestuous courtship². In the long-term, a vaccine or therapy will be needed and this is where animals showing some resistance to the tumours may be invaluable.

What's more, devils living in western Tasmania, which seem to be genetically distinct from their eastern cousins³, have so far remained free of the disease. Stephen Pyecroft, a veterinary pathologist at the DPIWE's animal-health laboratories in Mount Pleasant, is now scrutinizing the genes of devils from different regions to look for any natural genetic variation that may confer resistance.

But in the short term, the devils' best chance lies with the work of Menna Jones, a biologist at the DPIWE and the University of Tasmania in Hobart, who is investigating whether removing sick animals from wild populations suppresses the disease.

Jones has just completed a pilot study on two connected peninsulas on Tasmania's southeastern coast. These are ideal for restricting the movement of devils because they are separated from the mainland by a single bridge. She removed all infected animals from a 120-square-kilometre region in the peninsulas, and her unpublished results,

presented at the 2005 Australasian Wildlife Management Society conference in Hobart, concluded that the population structure remained intact and no new cases of the disease were reported.

"This work is the best hope for the devil yet," says Nick Mooney, a wildlife officer for the DPIWE. Andy Dobson, an ecologist studying infectious diseases at Princeton University, New Jersey, agrees: "The initial results are giving us grounds for optimism".

Total wipe-out

Jones is now conducting a larger two-year study to completely eradicate the disease from the peninsulas. "If successful, this will be the only disease-free reserve of Tasmanian devils," says Jones. They will safeguard the bridge against roaming devils and set up remote cameras across the region to detect any diseased animals that manage to avoid trapping.

However, it's not clear that this approach will work for all areas. "In a larger open system, where animals can come and go, you may be missing against the wind," says Mooney.

Others are optimistic that tackling the problem from a multitude of angles will ultimately save the devil. "You can be a pragmatic scientist about it," says Pyecroft, "but you also have to have a bit of faith."

Carina Dennis is Nature's Australasian correspondent.

1. Hawkins, C. E. et al. *Biol. Conserv.* (in press).

2. Pearce, A.-M. & Swift, K. *Nature* **439**, 549 (2006).

3. Jones, M. E., Paetkau, D., Geffen, E. & Moritz, C. *Mol. Ecol.* **13**, 2197-2209 (2004).

BUSINESS

Ticking the right boxes

The humble tick isn't often viewed as an ally of human health (see page 524), but one British biotechnology company is banking on its hidden charms.

Reading-based Evlutec holds patent rights on a number of proteins found in the saliva of ticks, and some are already showing clinical promise, particularly for the treatment of allergies and inflammation.

The proteins' potency derives from a tick's need to evade detection while it feeds on its animal host. Its saliva contains a number of molecules that suppress its victim's immune response, allowing the tick to feed for days unnoticed.

"It's the tick's stealth technology," says Mark Carnegie Brown, Evlutec's chief executive, adding that it's a technology that is yielding molecules with therapeutic potential.

The story began almost 20 years ago at a laboratory in Oxford run by the Natural Environment Research Council (NERC), when researchers led by virologist Patricia Nuttall started investigating how ticks use proteins in their saliva to suppress the immune system of their hosts.

Nuttall, who now directs the NERC's Centre for Ecology and Hydrology in Swindon, soon discovered an array of proteins of interest. "These molecules have been refined by millions of years of evolution," she explains. "There are no toxicity problems, they work on a range of animals, they aren't fragile — and there are an awful lot of them."

One of the most promising proteins, dubbed rEV131, binds to histamine, which, when produced by the body in excessive amounts, is associated with allergies and inflammation. Anti-inflammatory drugs tend to work by blocking one of at least four different histamine receptors in the body, but Nuttall describes rEV131's behaviour as "much more efficient". The tick protein grabs hold of the histamine itself and so stops it binding to the receptors.

The commercial potential seemed clear, and in 1998 Evlutec emerged as the first spin-off company from the NERC, backed by 3i, the London-based venture-capital group. The patents were held by the NERC, rather than

the scientists, and the company was set up separately from the research team — although Nuttall served as a non-executive director from 2000 to 2003, and is still advising the company on the possibility of using the proteins to vaccinate animals against ticks.

Last year, rEV131 was shown to be effective for treating hay fever in a second-stage clinical trial run by Paul Ratner, an allergist who runs Sylvana Research in San Antonio, Texas. Evlutec plans to run a second phase II trial this year to determine the right dose and frequency for the therapy. This stage will be critical

to the company's prospects for making it big, says Michael Aitkenhead, an analyst at Bridgewell Securities in London. Last September, Bridgewell organized Evlutec's second round of financing in 2005; between them, the two rounds raised nearly £20 million (US\$36 million) from institutional investors.

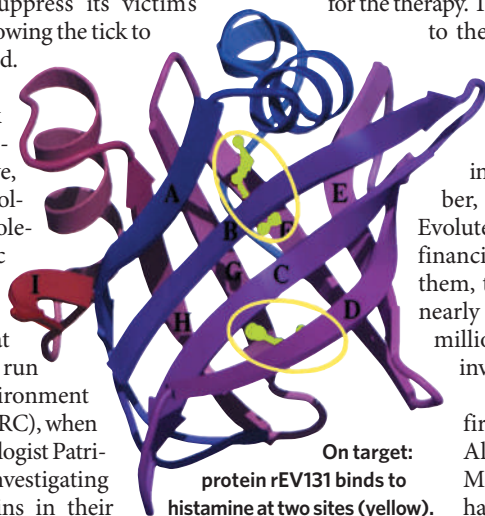
Evlutec, which was first listed on London's Alternative Investment Market in August 2004, has a small core of just 11 full-time staff. It subcon-

tracts out many operations including the running of clinical trials, and has enough cash to keep going for another 18 months as it searches for a partner to take rEV131 through the large, phase III trials needed for regulatory approval.

The tick protein is also in trials for suppressing inflammation after cataract surgery and for treating 'dry eye'. But it is only one of 16 proteins from tick saliva that Evlutec holds patents on or applied to patent; some show promise in animals for treating heart attacks and autoimmune diseases. The firm also has a partnership with animal-health company Merial of Duluth, Georgia, to use one of the proteins as a vaccine against both ticks and tick-borne diseases in cattle.

Carnegie Brown says he jumped at the chance to join the young company in 2003 "for the opportunity to build a business with some really fascinating technology". The company has enough funds to do at least three clinical trials in 2006, he says: "There's a whole series of strings to our bow."

Colin Macilwain



On target:
protein rEV131 binds to
histamine at two sites (yellow).

EVLUTEC

IN BRIEF

FEW OFFERS The number of initial public offerings for companies backed by venture-capital groups in the United States fell last year from 67 to just 41, reports VentureOne, a division of Dow Jones. The level of activity is down drastically not just from the peak level of 250 offerings in 1999, but from normal levels before then. Low investor interest and a regulatory clampdown on Wall Street are being blamed for the slide.

CELTIC TIGER Amgen, one of the world's largest biotechnology companies, has announced plans to build a US\$1-billion manufacturing plant in Cork, Ireland. The California company says that more than 1,100 people will be employed at the new plant, and that Ireland had been selected for its growing pharmaceutical industry and its low corporate taxes. It also plans to open a new clinical research facility at Uxbridge, near London.

NUCLEAR READY Toshiba is set to purchase Westinghouse, the world's largest builder of nuclear reactors, for about US\$5 billion. The acquisition is the largest-ever foreign purchase by the Japanese consumer electronics group. And the purchase price is almost double what was anticipated when UK-based BNFL put the Pittsburgh company up for sale last year. This reflects growing expectations that the construction of nuclear power plants is about to see a revival worldwide after decades in the doldrums.

UP FOR SALE Biotechnology company ImClone Systems of New York has appointed the investment bank Lazards to find it a partner or a buyer. At the same time, its acting chief executive is being replaced, for the second time in three months. The company, whose main product is the cancer drug Erbitux, has been in trouble since its former boss Samuel Waksal was jailed for insider trading in 2003. Analysts caution that, as a biotechnology company whose fortunes largely hinge on a single product that is facing increasingly stiff competition, ImClone may struggle to find a buyer.

Gene-function wiki would let biologists pool worldwide resources

SIR — Your Special Report “Internet encyclopaedias go head to head” (*Nature* 438, 900–901; 2005) shows that Wikipedia comes close to Britannica in terms of the accuracy of its science entries.

As a frequent user of Wikipedia and also a biologist, I hope that one day a wiki on gene function will be voluntarily created and maintained by biologists.

After a microarray experiment or a BLAST search, hundreds or thousands of interesting gene names are revealed, but the average biologist has no clue as to the function of most of them. Following up with a literature search wastes a lot of researchers' time and energy.

A wiki on gene function would make life very much easier for biologists. Such a wiki would also be less susceptible to spam, as most users would be biologists. It would also be more accurate, as long as statements are accompanied by references.

Several gene-function databases already exist, but each has certain disadvantages for the average biologist.

The Entrez Gene database (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene) contains a Gene Reference Into Function, or GeneRIF, that allows users to submit statements with supporting literature references to annotate gene function. But although many short statements have accumulated in this database, it would be useful to have concise summaries compiled by experts, and these are not usually available.

The Molecule Pages for signalling proteins (www.signaling-gateway.org/molecule) and the OMIM database for human inherited diseases (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) both rely heavily on community input, but are both limited to special-interest groups.

The UniProt Knowledge Base (www.uniprot.org) provides an integrated view of gene functions. But the information is collected from highly heterogeneous experimental and computational data sources, so the annotations often lack confidence measures.

In addition, the entries in these gene-function databases — unlike those in Wikipedia — cannot be conveniently commented on or edited by average biologists. This discourages input from many potential contributors. Concise and accurate gene-function annotations compiled and edited by human experts are preferable in many cases.

A wiki on gene function, which utilizes the collective brain power of biologists

around the world, would be an invaluable tool for biological sciences.

Kai Wang

Department of Microbiology,
University of Washington, Box 358070,
Seattle, Washington 98195, USA

Conferences that welcome spouses aid research too

SIR — In the struggle to balance career and family, long scientific conferences can represent a particular challenge. Occasionally, researchers may be tempted to invite their spouses. Not only is this a chance to spend time together, but it also provides an opportunity for the spouse to meet the key people in the researcher's field and put names together with faces.

This plan can present difficulties, as I recently discovered when I travelled with my wife and baby daughter to a Keystone conference my wife was attending in Utah. Although she encouraged us to join her for the social hour and dinner, we were told this daily event was “for scientists only”. However, we expressed concern about this policy, and by the end of the week, the Keystone website announced that spouses would now be permitted to attend the evening social hours: a pleasing response.

Given the all-consuming nature of scientific research, adopting a more inclusive policy towards interested spouses would be a small gesture by every conference organizer that could go a surprisingly long way towards improving the happiness of researchers — and of those who support them.

David A. Shaywitz

Department of Molecular and Cellular Biology,
Harvard, and Neuroendocrine Unit, Bulfinch 457,
Massachusetts General Hospital, 55 Fruit Street,
Boston, Massachusetts 02114, USA

Mudskippers undermine ID claims on macroevolution

SIR — In his Science in Culture article “Dying for a drink” (*Nature* 438, 564; 2005), Martin Kemp outlines the ancestral line of vertebrates, from modern humans back through evolutionary history to mudskippers.

Mudskippers are amphibious gobioids (a suborder of perch-related bony fishes) with arm-like pectoral fins. They live in the mangrove swamps of Africa, Asia and Australia, where they feed both in water and on land. The most terrestrial species catch insects and sometimes climb trees. They are living model organisms for the study of a key event in the history of life: the evolutionary transition of fishes to amphibians (tetrapods)

that occurred about 364 million years ago.

Modern young-Earth creationists and adherents of the intelligent design (ID) movement have no problem with microevolution (speciation). But most of these Bible-based anti-darwinists refuse to accept macroevolution (phylogenetic development above the species level) on the grounds both that it is unscriptural and that it has never been observed (S.B. Carroll *Nature* 409, 669; 2001).

Mudskipping gobies and other amphibious fishes are examples of macroevolution in progress that can be analysed by observation and experiment. They are living intermediate forms that display a number of anatomical and physiological macromodifications of their fishlike body plan that enable them to live and forage on land.

These facts are relevant to the current ID debate, as they illustrate Darwin's classical concept of descent with modification that evolved over past decades into the modern (synthetic) theory of biological evolution — the unifying principle of all life sciences.

U. Kutschera

Institute of Biology, University of Kassel,
Heinrich-Plett-Strasse 40,
D-34109 Kassel, Germany

Reader-appeal should not outweigh merit of research

SIR — There is one aspect of *Nature's* acceptance criteria that your Editorial on peer review (“Three cheers for peers” *Nature* 439, 118; 2006) does not consider.

The broad audience of *Nature* forces its editors to pre-screen papers according to how appealing they will be for its readers, even if appeal and importance do not always go hand in hand. This is absolutely legitimate, given the broad character of the journal, given its independence (it is a private enterprise, after all) and given the fact that any author can choose whether to submit papers to *Nature* or not.

But special consideration is due to the growing weight that authoring papers in *Nature* is acquiring in personal curricula. Gratifying though this must be for *Nature's* editors, it has the slightly worrying implication that bright young scientists are beginning to be driven more by the appeal of a potential paper than by its importance — a trend to which the scientific community should find a response.

The long and exemplary relationship between *Nature* and the scientific community allows me to hope that the journal itself will help in this endeavour.

Emilio Artacho

Department of Earth Sciences,
University of Cambridge, Downing Street,
Cambridge CB2 3EQ, UK

BOOKS & ARTS

A natural history of religion

A darwinian philosopher turns his attention to the strength of religion in the United States.

Breaking the Spell: Religion as a Natural Phenomenon

Daniel C. Dennett

Viking/Allen Lane: 2006. 488 pp. \$25.95.

To be published in the UK in March.

Michael Ruse

The US presidential election of 2004 showed that the nation is divided. On either coast the states are blue, a colour that, somewhat confusingly for Europeans, represents the political left. In the middle are the red states, which are on the right. But this is more than just a political divide. It is also very much a religious one, with the people of the red states being largely enthusiastic evangelicals; the blue states host a motley collection of episcopalians, unitarians and others, joined uneasily by non-believers of various shades.

It is against this background that we should read *Breaking the Spell* by the US philosopher Daniel Dennett, a notorious non-believer. You would not expect this book to bring comfort to Christians, and it certainly does not. So little impressed is Dennett by religion's claims to truth that he does not even bother to produce new material. He simply quotes at length from earlier writings. He is nevertheless trying to move the discussion forward, following in the tradition of the eighteenth-century Scottish philosopher David Hume and providing a natural history of religion. Dennett aims to give an account of how and why religion appeared, and how and why it has the hold it has today. As you might expect, given that he is an ardent darwinian, for Dennett religion and its origins are ultimately a matter of biological fitness to survive and conquer.

In the United States today there is a need for a good book on religion, and on how we might start to bridge the gap between red and blue. There is much that is pertinent to this end in Dennett's discussion. I was particularly interested in his coverage of the ways in which US mega-churches cater to people's needs and wants. In a sense, it is all a bit like toothpaste or automobiles — who has the product with the edge that captures the consumer's attention and opens his or her wallet? Do not bother with details such as the Trinity or the meaning of Grace — just make sure there is plenty of free parking and baby-sitting. There is a message here for those of us who think that all this would be farcical were it not for the fact that people

who attend such institutions tend to hold moral and social values that lead to anti-abortion fanaticism, capital punishment, excoriation of gays and lesbians, and dangerous military excursions in the Near East. If we do not like what the churches are feeding people, we had better come up with an attractive alternative.

Ultimately, taken as a whole, Dennett's is not the book for which we search. He does give some prescriptions for action, generally (and admirably) involving education. But basically there is something off-key about the whole discussion. Part of the problem is philosophical. A major plank in Dennett's discussion is that religion is all smoke and mirrors, so even if we cannot hope immediately to eliminate all religious belief, those of us in the know will realize that we are dealing with a delusion, rather than a rationally justified belief system. However, a naturalistic analysis of religion in itself has no direct bearing on the truth of religious claims. My eyes are the end products of a long process of natural selection. Does that make any less real the truck I see bearing down on me as I stand in the middle of the road?

Most problematic is Dennett's blind spot regarding history. There is no real account of the way religion has developed and of how we have ended up where we are today. Another major weakness is the exclusive focus on the United States, which is a peculiar country where

religion plays a huge role, far bigger than in most of Europe. This difference is reflected in many diverse ways, particularly in the social values mentioned above. You cannot begin to talk about biological bases for religion — 'genes for God' and that sort of thing — without taking account of the fact that peoples of very similar biological background behave in very different ways about religion and its implications. Only history — the fact that the United States was founded by people with major religious concerns, and that this has persisted for four centuries — can help us to tease apart the cultural and the biological.

Dennett asks whether only a true believer can report properly on religion. He argues strongly that on a topic this important we all can, and should, engage. I agree. However, a degree of empathy is needed, and it is this that is missing. Unless you have some sense of what fires people up, you are never going to reach them or have any hope of shifting their beliefs. The debate over religion in the United States is intense and profoundly affects the status of science. I hardly have to remind *Nature* readers of the battle to introduce 'intelligent design' into biology classrooms. But we need better books than this to address the issues. ■

Michael Ruse is in the Department of Philosophy, Florida State University, Tallahassee, Florida 32306, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Religious leader: President George W. Bush begins cabinet meetings with prayers.

B. KRAFT/CORBIS

History by numbers

God Created the Integers: The Mathematical Breakthroughs that Changed History

edited by Stephen Hawking

Running Press: 2005. 1,160 pp. \$29.95

Jeremy Gray

The subtitle — the mathematical breakthroughs that changed history — is not mere advertising. It is a yardstick, in fact various yardsticks, for assessing the book as a whole. Stephen Hawking brings us generous selections of Euclid's *Elements*, the works of Archimedes and Diophantus, René Descartes' *Geometry* entire, Pierre-Simon Laplace's *A Philosophical Essay on Probability*, and material from Joseph Fourier, Carl Friedrich Gauss, Augustin-Louis Cauchy, Bernhard Riemann, Georg Cantor and Henri-Léon Lebesgue. There are also the first 11 chapters of George Boole's *An Investigation into the Laws of Thought*, Richard Dedekind's *Essays on the Theory of Numbers*, smaller extracts from Newton's *Principia*, and papers by Karl Weierstrass, Kurt Gödel and Alan Turing. All these have some claim to have changed history,

except perhaps the work by Diophantus, remarkable though it is. The same claim can be made of quite a few other works too, but as this book has nearly 1,200 pages it would be churlish to complain.

The items are well chosen. They are an interesting mix of the well known and the unexpected, and cover a range of topics in mathematics from geometry to mathematical analysis, probability and the modern foundations of mathematics. They range over an extensive period and, because many are presented either whole or at least extensively, they can be read with pleasure.

Every anthology faces the issue of what to do with the best-known pieces. Include them and some complain that they are too well known; omit them and others lament their loss. If this book is the only collection of original works of mathematics in translation that a reader owns, there is a lot to be said for it. But as an addition to the small but useful number of collections in English it is more annoying, because the opportunity was not taken to translate more works. All but four of the items here are already available in

English. The new ones are the work by Cauchy (some of which already exists in English elsewhere), some of the papers by Riemann (one of which is already in English), the passage by Weierstrass, and the extracts from the work of Lebesgue.

It is one thing to use an old translation if there is no significant improvement to be had in making a new one, but it is quite another to perpetuate an inadequate work. Here the Diophantus is annoying because it is the version by Thomas Little Heath. He turned the work into a densely written but by now old-fashioned school problem book, and the commentary, given in extensive footnotes, moves the reader even further from Diophantus' habit of mind. It is true that Diophantus' example is so strange that it reminds us how little we know about the ancient Greeks, but we do know how it was received in the Arab world and in the modern West, and it can be said to be only one of several works that helped bring about the 'modern algebra' of the seventeenth century. It would have been better if Hawking had included instead al-Khwarizmi's work on *al-jabr*, from which algebra takes

A culture of knowledge

An exhibition in Paris explores the golden age of Islamic science.

Pete Jeffs

There is a mosque near my studio in east Paris. Passing by at the hour of prayer, I find myself intrigued by an Islamic culture that dates from the eighth century, occupies a central place in world culture, and yet remains a mystery for other societies.

The exhibition 'The Golden Age of Arabic Sciences', which can be seen at the Institute of the Arab World in Paris until 19 March, sheds light on a people who produced exquisite manuscripts, developed experimental sciences, and added new disciplines to those of the classical world.

The exhibition questions the idea that Arab-Islamic science simply translated key classical sources of knowledge and passed them on to the West. It shows how the Arabs of the expanding empire transformed and extended Greek, Mesopotamian, Persian and Indian ideas, and turned theory into practice. Analysis preceded assimilation.

The optics of Ibn al-Haytham (AD 965–1040), for example, encouraged an experimental approach to the pursuit of physics. In 830, al-Khwarizmi, the acknowledged founder of algebra, detailed the necessary mathematics for calculating inheritances, conducting

commerce and constructing canals. Baconians would surely have approved.

By this time, paper was becoming the preferred vector for the transmission of Arabic science. Industrial-scale production began around 750, first in Samarkand and then in Baghdad. Cheap paper rapidly outstripped the use of papyrus and parchment. Almost 4 million individual Arabic writings from the eighth century onwards are known to have been conserved by libraries throughout the world.

Arab society integrated its knowledge by building hospitals, observatories and libraries, beginning with the House of Wisdom — Bayt al-Hikma — in Baghdad, shortly after 800.

Translations and original writings were stored in more than a thousand different public and semi-public libraries scattered throughout the empire, catering for a widespread demand for scholarship. Knowledge was transmitted by scholars who

travelled regularly, as well as through written correspondence. The Koran, incidentally, encouraged scientific activity.

Some 200 manuscripts and objects have been assembled in Paris for the exhibition, including this planispheric astrolabe from the Nasser D. Khalili Collection and a profusely illustrated copy of Dioscorides' *Herbal*. But the exhibition opens with a map. When we

can associate cities such as Damascus or Maragheh with astronomy as readily as we associate Padua with anatomy or Leiden with electricity, Arab-Islamic culture will be on its way to a fuller appreciation in the West. This exhibition offers a fine lesson in the meaning of the word 'civilization'.

Pete Jeffs is an artist based in Paris, France.



NASSER D. KHALILI COLLECTION

its name (see 'A culture of knowledge').

As for Euclid and Archimedes, there are 240 pages of Heath's editions with their erudite but dated commentaries. These may well help the reader with the mathematics, but historical scholarship has moved on. Unfortunately, the new comments are not much better. Each item is introduced with remarks that vary from the personal and insightful to the tired and incorrect. Among the latter is the supposed Greek crisis of the incommensurables, which was once presumed to have derailed the

pythagoreans and which most historians these days think had little effect. The generally more accurate account of the life and work of Archimedes fails to mention the fact that the only manuscript of *The Method* — the most interesting of his works, in which he explained how he came to his discoveries — has recently re-entered the public domain after disappearing for most of the twentieth century.

Similar comments could be made about the more modern entries. These too are generally accurate but the origin of the information is

not stated, so readers have no chance to catch up with contemporary scholarship in the history of mathematics. Nor can they find out how to sustain the flame of interest this book surely hopes to kindle, which is a pity, because Hawking's comments have an infectious enthusiasm for their subject and the book contains some great works. ■

Jeremy Gray is at the Centre for the History of the Mathematical Sciences, Faculty of Mathematics and Computing, Open University, Milton Keynes MK7 6AA, UK.

Sex and power

Sexual Conflict

by Göran Arnqvist & Locke Rowe

Princeton University Press: 2005. 360 pp.
\$99.50, £55 (hbk); \$39.50, £26.95 (pbk)

Tracey Chapman

Sexual conflict occurs because of the different evolutionary interests of males and females in reproductive decisions. Indeed, the differential investment of males and females in reproductive episodes and the low relatedness of mating partners make sexual conflict almost inevitable. Geoff Parker clearly laid out the theoretical basis for sexual conflict in the 1970s, but it is only over the past ten or so years that a significant body of empirical work has started to emerge.

There has, however, been little agreement about what sexual conflict is, or what might constitute unambiguous evidence for it. In addition, the relative importance of sexual conflict in driving evolutionary change, and the extent to which it could contribute to reproductive isolation and speciation, are unknown. There is also considerable confusion about precisely how coevolution driven by sexual conflict is distinct from traditional models of coevolutionary change by sexual selection (that which arises from competition between individuals of the same sex for matings, or from mate choice).

Into this uncertainty comes the excellent and wide-ranging *Sexual Conflict* by Göran Arnqvist and Locke Rowe, the first book-length treatment of this emerging field. The authors have done a great service in defining the field and in illuminating some of its conceptual difficulties. They also bring together a rich treasury of examples that must surely stimulate discussion and further study.

The greatest strength of the book is in tackling the theory. Arnqvist and Rowe do a first-rate job of dissecting models of sexual selection and sexual conflict, and in getting to grips with the distinctions between them. The book is well worth reading for this alone. For example, the reader is led clearly through the idea that coevolution driven by sexual conflict is distinct from sexual selection: under sexual

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sexual partners often seem to be pulling in different directions.

conflict, the selection on females is direct and focused on reducing female mating costs.

Such costs cannot be incorporated into the traditional models of sexual selection developed by R. A. Fisher in the 1930s, which propose a genetic benefit of female mate choice through the enhanced mating ability of sons. In contrast, costs of choice are not a problem for 'good genes' (indicator) models, but these assume that female choice will result in the acquisition of genetic fitness benefits for the offspring, and this is not generally assumed to occur under sexual conflict. The non-adaptive result of female mating biases also distinguishes sexual conflict from direct-benefit models.

Sexual conflict could therefore operate within this traditional sexual-selection framework but be distinguished from it by the nature of the forces acting on female mating biases. However, the authors also suggest that these standard equilibrium models may be inappropriate for analysing sexual conflict. For example, if males had a vast array of ways in which to manipulate females, and females had a similarly large number of ways in which to respond, the dynamics of this contest could well be more suited to non-equilibrium modelling. If this were true, then evolutionary change driven by sexual conflict would be different from any processes of mate choice with which we are familiar.

that a process of continual adaptation and counter-adaptation could obscure evolutionary change resulting from sexual conflict. It may also be difficult to measure the costs and benefits of reproductive decisions under an appropriate range of relevant conditions, or even to decide what data are required. As the authors state, "unambiguous experimental data of sexual conflict are really quite scarce".

The best part of the ensuing discussion of the empirical studies in which sexual conflict could be operating covers the species for which good data are available, notably bed bugs, diving beetles, water-striders, fruitflies and dunnocks. The authors also consider numerous further cases in which sexual conflict might be occurring. These are more anecdotal but provide some fascinating natural history, such as the use of gin traps by male sage bush crickets, the 'playing dead' strategy of some species of robber fly, and the male funnel-web spider, which drugs its mate before copulation. This remarkable wealth of examples hints at the potentially ubiquitous distribution and importance of sexual conflict, and represents an extremely valuable resource that should stimulate further study and experimentation on the systems described. ■

Tracey Chapman is in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

The third Bond

When James Watson and Francis Crick unveiled their structure of DNA, one of the two kinds of base pair in the molecule was given two hydrogen bonds instead of three. Who spotted the third bond and when?

Simon Wain-Hobson

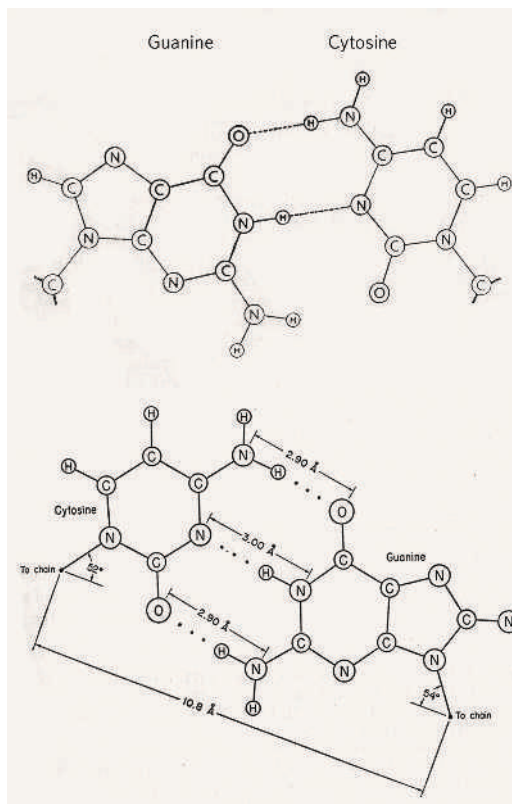
It is a truth universally acknowledged that a guanine–cytosine (GC) base pair has three hydrogen bonds whereas adenine–thymine (AT) has two. But James Watson and Francis Crick didn't see it that way back in 1953 when they published the structure of DNA. In their second DNA paper published in May of that year, the GC base pair is shown with only two hydrogen bonds (see top figure). So who spotted the third bond? When? And why was it initially passed over?

In his book *The Double Helix*, Watson notes that "The formation of a third hydrogen bond between guanine and cytosine was considered but rejected because a crystallographic study of guanine hinted that it would be very weak".

The third hydrogen bond in a GC pair makes its first published appearance in a paper by Linus Pauling and Robert Corey¹ in 1956 (see bottom figure). However, the first hint of the third bond in the scientific literature actually comes in a footnote to a paper published earlier that year by Jerry Donohue, a physical chemist and crystallographer.

In this paper², which describes the possible ways in which pyrimidines and purines might hydrogen bond to one another, Donohue notes, "It has been pointed out by Professor Pauling that it is possible with only small distortion for guanine and cytosine to pair by formation of three hydrogen bonds... The formation of this additional hydrogen bond may confer extra stability on the Watson–Crick Structure." The respectful tone is understandable given that Pauling recommended Donohue's paper to the *Proceedings of the National Academy of Sciences* on 23 November, 1955. So Pauling had the third bond by the end of that year.

But what was the guanine crystal structure alluded to in *The Double Helix* that led Watson and Crick to reject the third bond? Their colleagues at the Cavendish Laboratory in Cambridge, under the direction of Lawrence Bragg, had been working on the structure of pyrimidines, purines and nucleosides since 1948, including adenine, guanine hydrochloride and a uracil derivative. Meanwhile, down in Birkbeck College, London, another group had published the structure of cytidine. So by spring 1953 initial structures of the four bases were either known or could be reasonably



The third hydrogen bond in a guanine–cytosine base pair (bottom) was missed in the 1953 description of DNA (top).

inferred. These data would have been available to Watson and Crick.

The adenine and guanine structures used in Watson and Crick's figure seem to be those determined by Bill Cochran and June Broomhead of the Cavendish Laboratory. Telltale signs are in the guanine structure — the bonds surrounding the keto and amino groups are irregular, distorting this part of the structure.

In Watson and Crick's figure, the hydrogen-donating amino group in the guanine base leans away from the keto acceptor group of cytidine (see top figure). Hydrogen bonds are at their strongest when the hydrogen atom and the donor and acceptor atoms are aligned linearly. Any third bond drawn on this figure would be at best weak with a 'kink' of about 18° from this linear position, and would have been a little on the long side at 3.3 angstroms.

But why did Watson and Crick reject even a weak third bond? The answer may lie back in Donohue's 1956 paper². In that paper on hydrogen-bonding patterns between purines and pyrimidines, "a maximum deviation of N–H...X from linearity

of about 15° was allowed".

Donohue shared the same office as Watson and Crick at the Cavendish Laboratory. It was he who advised Watson over which tautomeric forms of pyrimidines and purines to use in their DNA model.

The acknowledgement, "We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially in inter-atomic distances," appears at the end of the first DNA paper — indeed before mention of Maurice Wilkins and Rosalind Franklin, both key players in the discovery of DNA's structure. So it may be presumed that Watson and Crick deferred to Donohue and cut the third bond.

Pauling and Corey, however, arrived at the right structure thanks to a strong dose of structural common sense. They note that the structure for guanine contains "a small error" in that angles of the bonds adjacent to the keto group are irregular.

While working from the literature, they made many "reasonable arguments based upon considerations of electronic structure", one of which was that equal bond angles surround the keto and amino groups.

Using a "reasonable" structure for guanine, the third bond falls into place like a charm. Indeed, the third bond proved to be every bit as good as any of the other hydrogen bonds in AT and GC pairs coming in at 2.9 angstroms, the N–H...O hydrogen bond being essentially linear. The bottom line is that there is a trace of Pauling in the double helix.

1953 was an excellent year — the structure of DNA, the Miller–Urey experiment, and the death of Stalin. And of course with Casino Royale the other Bond, James Bond, first stepped off the page in 1953.

Simon Wain-Hobson is at the Pasteur Institute, Paris.

FURTHER READING

1. Pauling, L. & Corey, R. B. *Arch. Biochem. Biophys.* **65**, 164–181 (1956).
2. Donohue, J. *Proc. Natl. Acad. Sci. USA* **42**, 60–65 (1956).

ACADEMIC PRESS

ESSAY

NEWS & VIEWS

SOLAR SYSTEM

A planet more, a planet less?

Scott S. Sheppard

Further observations of an object dubbed 2003 UB₃₁₃, which lies beyond Neptune, show that its diameter is around 3,100 kilometres. This makes it larger than Pluto, the smallest 'traditional' Solar System planet.

The number of objects detected in the Solar System beyond the orbit of Neptune — so-called trans-neptunian objects — is increasing rapidly. This is mainly due to large, sensitive digital detectors on telescopes and sophisticated software, running on high-speed computers, that can detect moving objects. More than 1,000 of these objects have been discovered since the first one — other than Pluto — was detected in 1992. On page 563 of this issue, Bertoldi *et al.*¹ describe observations of the brightest such object, known as 2003 UB₃₁₃. They find that it has a diameter of about 3,100 kilometres, making it the largest Solar System object discovered since Neptune, and the first known trans-neptunian object larger than Pluto (Fig. 1).

Determining the size of distant objects such as 2003 UB₃₁₃, which is 97 times farther from the Sun than is Earth, is not straightforward. Pluto was discovered in 1930, but it took decades to get a handle on its diameter: initial estimates put its size at similar to that of Earth, but now it seems that it is smaller than the Moon. As 2003 UB₃₁₃ is about three times farther from the Sun than is Pluto, modern technology cannot resolve it: it appears as just a point of light even in images of the highest resolution.

There are, however, indirect methods for

measuring the size of an unresolved object². When sunlight strikes the surface of an object, a fraction of the incident light is reflected back into space. This fraction is known as the albedo of an object and can be observed in the visible region of the electromagnetic spectrum. An object is therefore optically bright either simply because it has a large diameter (and so more surface area from which to reflect light), or because it has a high albedo. To determine the albedo and diameter of an object separately, we need to determine what proportion of the incident light it reflects and what proportion it absorbs. This can be done by measuring the object's thermal radiation, because absorbed incident light warms a body's surface and is re-radiated back into space as heat. At its distance from the Sun, 2003 UB₃₁₃ should receive enough sunlight to have a temperature of about 25 kelvin (−248 °C), the exact value depending on its albedo. At that temperature, a perfectly absorbing and emitting object would emit peak 'black body' radiation at a radio wavelength of around 0.1 millimetre.

Bertoldi and colleagues¹ used the radio telescope operated by the Institute for Millimetre Radio Astronomy (IRAM) in the Sierra Nevada mountain range of southern Spain to make observations at the requisite wavelength, and thus detect for the first time the thermal

radiation of 2003 UB₃₁₃. Combining these observations with observations at visible wavelengths³, they find that 2003 UB₃₁₃ has a very high albedo, with about 60% of the sunlight that strikes it being reflected back into space. The authors also determined that the diameter of 2003 UB₃₁₃ is 3,100 kilometres (to within 300 km). Pluto's diameter, by comparison, is around 2,300 km.

The diameters and albedos of only about a dozen trans-neptunian objects have so far been measured accurately. Most of these results stem from thermal measurements at millimetre and submillimetre radio wavelengths, or still smaller far-infrared wavelengths, using NASA's Spitzer Space Telescope. The albedos measured range from around 3% to 60%, possibly with slight systematic variations because of modelling uncertainties in the degree to which the objects' surfaces emit thermal radiation in the different wavelength regimes^{4–6}. Although the data are sparse, it seems that the largest objects have the highest albedos. This could be because gravity on these objects is large enough for them to have active atmospheres and so be able to retain volatile gaseous substances that could brighten their surfaces.

The trans-neptunian object 2003 UB₃₁₃ is not unique. It is typical of many of the icy



Figure 1 | Size matters. 2003 UB₃₁₃ is currently the fifteenth-largest known Solar System object, with a diameter larger than those of Pluto and the neptunian moon Triton, but smaller than those of Earth's Moon and Titan, the largest of Saturn's satellites. Other objects, besides the eight undisputed major planets, that are larger than 2003 UB₃₁₃ are Jupiter's satellites Callisto,

Io, Europa (not pictured) and Ganymede (the largest moon in the Solar System). Behind Pluto, the next-largest known trans-neptunian objects are 2005 FY₉, with a diameter of about 1,800 km, and Sedna (1,700 km). A handful more are larger than Ceres, the largest asteroid in the main asteroid belt between Mars and Jupiter.

planetoids that populate this region of the Solar System — known as the scattered Kuiper disk — in that its orbit is highly elliptical, with a distance of closest approach to the Sun (perihelion) near the orbit of Neptune. This type of orbit is unlike that of Pluto or other objects in the nearer, classical Kuiper belt, but is common to many trans-neptunian objects. Such bodies keep to less-elliptical orbits just beyond Neptune, lying 35 to 50 times farther from the Sun than Earth is. Like Pluto, 2003 UB₃₁₃ has a satellite. Physical observations³ of the surface of 2003 UB₃₁₃ have also shown it to be remarkably similar both to Pluto and to Neptune's satellite Triton (thought to be a captured Kuiper-belt object), both of which have high albedos and methane ice on their surface.

These observations lead inevitably to the question of whether 2003 UB₃₁₃ itself qualifies as a planet. The minimum size required for this has never been strictly defined. Until the nineteenth century, the term 'planet' was uncontroversial: a planet was any object that had a fairly circular orbit around the Sun, and that did not show any cometary activity such as a coma or a tail. By 1807, the word encompassed Ceres, Pallas, Juno and Vesta, which are in the belt of objects between Mars and Jupiter. In the mid-nineteenth century, the growing number of such objects being discovered caused the term 'asteroid' or 'minor planet' to be coined. A similar situation is now looming with Pluto, 2003 UB₃₁₃ and the rest of the trans-neptunian objects. Although the largest of these are much bigger than the largest main-belt asteroids, they also form an ensemble of bodies of probably very similar formation and evolutionary history.

Whichever way you care to count them, with the discovery and measurement of the size of 2003 UB₃₁₃ there are no longer nine major planets in the Solar System. A committee has been formed by the International Astronomical Union to mull over several competing definitions of the term. One fairly simple solution is that anything orbiting the Sun and larger than a certain minimum size counts as a planet. The size is likely to be arbitrary: a round-number radius of 1,000 km, or perhaps the size of Pluto itself. These bounds stem from a perceived need to maintain Pluto's 'traditional' status as a planet, and would leave us with ten current planets, including 2003 UB₃₁₃.

A second possible definition is that anything with a high enough mass to be spherical is a planet. Although this is a physical criterion, determining whether or not an object meets it is complicated. It would also increase the number of known planets by several factors. A third alternative is that any object that has a unique orbit (which means that it is sufficiently far removed from other orbiting bodies), and that dominates its local environment gravitationally, should be known as a planet. This definition has the most solid scientific basis, as objects with similar formation and evolutionary histories would be grouped

together. It would leave us with eight planets: Pluto would not fulfil the criteria.

A final proposal for what constitutes a planet takes into account the fact that the currently known major planets are themselves remarkably different, and can be split into sub-categories. Mercury, Venus, Earth and Mars, for example, are terrestrial planets whose compositions are dominated by rock. Jupiter and Saturn are gas giant planets dominated by their hydrogen and helium envelopes. Uranus and Neptune are ice giant planets, dominated by gases other than hydrogen and helium. The trans-neptunian objects, or ice dwarf planets, are probably composed of large amounts of volatile solids such as methane ice and water ice.

So 'planet' is a term, seemingly, that can be bent to any number of uses. But whatever the planet-definition committee decides will be

secondary to the naked fact of our rapidly expanding scientific knowledge and understanding of the Solar System. With further improvements in instrumentation and software, there may be further surprises in store. ■
Scott S. Sheppard is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015, USA.
e-mail: sheppard@dtm.ciw.edu

1. Bertoldi, F., Altenhoff, W., Weiss, A., Menten, K. M. & Thum, C. *Nature* **439**, 563–564 (2006).
2. Jewitt, D., Aussel, H. & Evans, A. *Nature* **411**, 446–447 (2001).
3. Brown, M., Trujillo, C. & Rabinowitz, D. *Astrophys. J.* **635**, L97–L100 (2005).
4. Altenhoff, W., Bertoldi, F. & Menten, K. *Astron. Astrophys.* **415**, 771–775 (2004).
5. Grundy, W., Noll, K. & Stephens, D. *Icarus* **176**, 184–191 (2005).
6. Cruikshank, D. P. et al. *Astrophys. J.* **624**, L53–L56 (2005).

MOLECULAR BIOLOGY

Prime-time progress

Stephen D. Bell

DNA is duplicated within a complex macromolecular machine. Insights into how replication begins and how this is coordinated with progression of DNA synthesis come from a diverse range of sources.

DNA is replicated by unzipping the double helix to expose the bases that act as a template for copying the genetic material. Both strands of DNA serve as templates, and thus one double helix becomes two. Conceptually, this is a simple reaction, but the devil — as so often — is in the detail: the process is mediated by a multitude of proteins and turns out to be mechanically complex. A trio of papers in this issue^{1–3} have made considerable headway in understanding the intricacies of replication.

One level of complexity in the replication reaction comes from the fact that DNA polymerase, the enzyme that synthesizes the new DNA, cannot begin a strand itself. Rather, it extends a short RNA 'primer' that is already bound to the template. This means that an additional enzyme, a primase, is required to generate the primer to start the polymerase reaction^{4,5}.

A second complexity lies in the fact that the two strands of the DNA double helix are arranged antiparallel to one another; the opposite directions are termed 5' to 3' and 3' to 5' (from the positions of carbon atoms in the sugars

that make up the DNA backbone). However, DNA polymerase can synthesize DNA in only one direction: 5' to 3' (Fig. 1). So the 3' to 5' template strand — the 'leading' strand — can readily be replicated, but how is the other, 'lagging' strand copied? This dilemma was resolved by the discovery that the lagging strand is replicated discontinuously. It is synthesized in short pieces, called Okazaki fragments, that are then joined together — essentially, the polymerase takes two 'steps' forward and then synthesizes one back. The

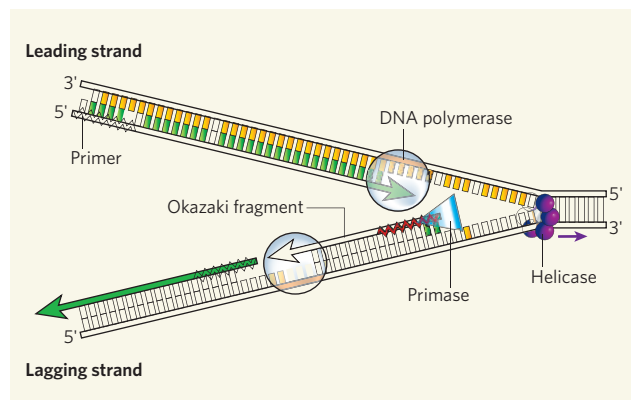


Figure 1 | DNA replication. The helicase unzips the double-stranded DNA for replication, making a forked structure. The primase generates short strands of RNA that bind to the single-stranded DNA to initiate DNA synthesis by the DNA polymerase. This enzyme can work only in the 5' to 3' direction, so it replicates the leading strand continuously. Lagging-strand replication is discontinuous, with short Okazaki fragments being formed that are later linked together.

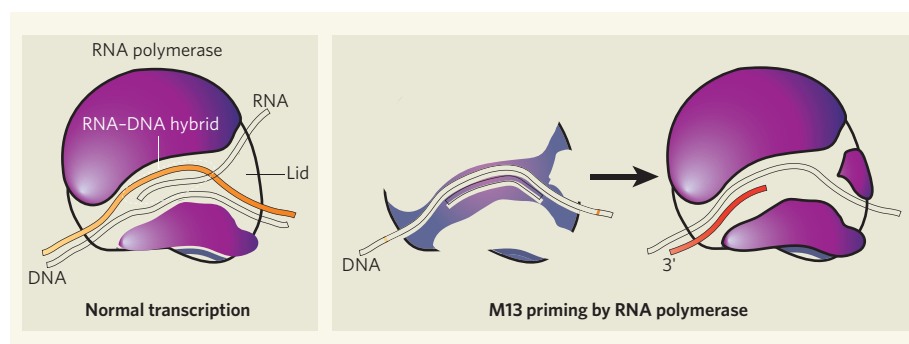


Figure 2 | Co-opting other enzymes for priming. The bacterial virus bacteriophage M13 replicates by using its host's RNA polymerase (which usually makes the RNA encoded in genes) to make RNA primers. Severinov *et al.*³ find that the single-stranded DNA that makes up the viral genome subverts the usual mechanism of the polymerase, which normally deals with double-stranded DNA.

difference in the synthesis of the two strands means that, in principle, the leading strand requires only a single priming event, whereas the lagging strand needs a new primer for each Okazaki fragment. And all these events must somehow be coordinated to produce two daughter strands at roughly the same rate.

One of the outstanding conundrums concerning replication is how the process deals with broken or damaged DNA, for example that generated by ultraviolet radiation, particularly at the leading strand. Replicating the damage could obviously be harmful for the daughter cells. So does the replication machinery just stall and wait for the damage to be fixed, or is there some way to sidestep the damage and continue without copying the damaged DNA?

Heller and Marians (page 557)¹ addressed this issue by biochemically mimicking a blockage of the leading-strand template. Their findings indicate that a new priming event can occur downstream of the lesion, so that the portion of the leading strand encompassing the lesion is not copied and the resulting DNA will be single-stranded. This would seem to contradict the prevalent theory that leading-strand synthesis is continuous. However, precisely this phenomenon is observed in bacterial cells, where ultraviolet irradiation can lead to single-stranded gaps on both leading and lagging strands. The gap is filled in afterwards using the remaining undamaged strand as a template. So presumably the existence of a mechanism for re-priming the leading strand by generating a single-strand gap buys the cell additional time to deal with the damage without interrupting the essential process of DNA replication.

How is the primase delivered to the site where it needs to act? In bacteria, the primase interacts physically with the enzyme responsible for unzipping the double helix, the replicative helicase. Some bacterial viruses that have their own replication machinery take this a step further by having the primase covalently attached to the helicase. This union of the unzipping and priming activities provides a mechanism to couple progression of the replication 'fork' with depositing the primase

on the lagging strand. Heller and Marians¹ reveal that the helicase is also crucial in re-priming the leading strand when replication is restarted after DNA damage.

The coupling of helicase and primase presents a potential dilemma. When the primase is deposited on the lagging strand, it synthesizes RNA in the opposite direction to the progression of the fork (Fig. 1). What happens to the motion of the fork when this happens? Do the enzymes uncouple, does the helicase run ahead, or does the fork stall transiently?

Lee *et al.* (page 621)² examine the priming process using a model DNA replication system derived from a bacterial virus called bacteriophage T7. This well-known system contains a covalently fused primase-helicase called gp4. The authors analyse the priming process using a single-molecule approach with a model DNA substrate and three purified proteins: gp4, the T7 DNA polymerase and an accessory factor, thioredoxin.

They find that the rate of leading-strand synthesis in the absence of lagging-strand synthesis is constant and uninterrupted. However, when ribonucleotides are added to the reaction, permitting the primase to act on the lagging-strand template, the progression of the leading strand pauses briefly at defined positions. These positions correlate well with the preferred priming sites of the gp4 primase, and each pause lasts about five seconds, in good agreement with the RNA synthesis rate of gp4 primase. However, in five seconds the leading-strand DNA polymerase can synthesize about 1,000 bases of DNA. So pausing the machinery prevents an uncoupling of the synthesis of the two strands. The stalling mechanism remains unknown, although Lee *et al.* speculate that the primase domain may regulate the DNA-unzipping activity of gp4. Although T7 gp4 has coupled primase and helicase domains in one protein, the physical interaction of primase and helicase seen in bacteria makes it likely that this pausing might occur in a broad range of organisms, even if it requires two proteins.

Does all priming have to be performed by a specialized primase? Some clues may come from the unusual replication process carried

out by 'selfish DNA' elements such as bacteriophage. Some of these elements have their own DNA replication machinery (such as phage T7, discussed above); but others, such as bacteriophage M13, co-opt the host cell's proteins to ensure their own replication. In fact, M13 subverts the host gene-expression machinery to initiate replication of its genome. The bacteriophage uses the bacterial RNA polymerase — usually employed in copying DNA into RNA destined to make protein — to produce the RNA primer for its replication instead. Normally, RNA polymerase peels the RNA off DNA, acting like a wire stripper, resulting in free single-stranded RNA and re-forming double-stranded DNA. But, to serve as a primer for M13 replication, the RNA produced by the polymerase must remain paired to DNA.

Severinov and colleagues (page 617)³ have uncovered the basis of this altered behaviour. The key lies in the fact that the M13 template is a single-stranded DNA. RNA polymerase usually acts on double-stranded DNA, which enters and leaves the enzyme molecule by specific channels, with only a short region unwound in the centre of the enzyme (Fig. 2). As the two DNA strands bind back together, a lid-like structure peels the RNA off and directs it out through a separate exit channel. But, because M13 is single stranded, there is no return to double-stranded DNA and the RNA-DNA hybrid becomes hyper-extended. This seems to force the hybrid molecule to reposition in the downstream channel of the RNA polymerase, with RNA extending beyond the body of the enzyme ready to act as a primer for DNA synthesis (Fig. 2). The authors propose that this may be a general property of the replication of a number of selfish elements in bacteria. Notably, several plant and animal viruses are proposed to extrude single-stranded regions at their DNA replication start sites, so perhaps RNA polymerase primes replication in some of these situations too.

Thus, bacteria and their viruses employ a range of strategies to effect priming, to couple priming events to the coordinated progression of the replication fork, and to ensure that DNA damage need not be an absolute barrier to fork progression. The power of these reconstituted bacterial and bacteriophage systems is apparent in the exquisite mechanistic detail that can be gleaned from their analysis. Similarly detailed studies of the more complex primases found in higher organisms⁵ are eagerly anticipated.

Stephen D. Bell is in the MRC Cancer Cell Unit, Hutchison MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK.

e-mail: sb419@hutchison-mrc.cam.ac.uk

- Heller, R. C. & Marians, K. J. *Nature* **439**, 557–562 (2006).
- Lee, J.-B. *et al.* *Nature* **439**, 621–624 (2006).
- Zenkin, N., Naryshkina, T., Kuznedelov, K. & Severinov, K. *Nature* **439**, 617–620 (2006).
- Kornberg, A. & Baker, T. A. *DNA Replication* 2nd edn (Freeman, New York, 1992).
- Frick, D. N. & Richardson, C. C. *Annu. Rev. Biochem.* **70**, 39–80 (2001).

MATERIALS SCIENCE

Colloids get complex

Alfons van Blaaderen

Self-organization of soft-matter components can create complex and beautiful structures. But the intricate structures created by adding a second stage of organization could reveal more than just a pretty face.

The term 'soft matter' denotes materials that are easily deformed by external stresses, and encompasses liquid crystals, polymers, surfactants and colloids (particles dispersed within another medium). Their basic constituents have characteristic sizes of between several nanometres and several micrometres, and, crucially, have the potential to self-organize, forming beautiful, regular three-dimensional structures. A triplet of recent papers^{1–3} presents the latest such structures: complex colloids formed through self-organization on scales up to a micrometre.

Alternative terms that have been used to describe these colloid structures — 'colloidal molecules', or 'patchy particles' — hardly do justice to their intricacy. What is considered a complex colloid is, admittedly, somewhat arbitrary: the colloidal 'ice-cream cones' (Fig. 1a) produced some years ago⁴, which resulted from repeated polymerization and the subsequent phase separation of the polymers formed, would certainly have merited the term complex colloid. The innovation of recent efforts, however, is that structures are being designed with a second stage of self-organization in mind. Such an approach, in which colloidal

particles are first formed at soft-matter scales, and then built up to far more intricate structures, should allow unprecedented control over the three-dimensional organization of materials, as well as the combination of different materials over several length scales.

The results of Cho *et al.*¹, published in the *Journal of the American Chemical Society*, exemplify the fruits of this technique. The authors created complex colloidal structures (Fig. 1b, c) by drying emulsion droplets containing 'bidisperse' charged colloids, consisting of components of two quite different sizes — one on the nanometre and one on the micrometre scale. Using the same or opposite charges on the two components, an amazing richness of structural motifs could be obtained. Equally impressive results have been published by Lin *et al.*² (Fig. 1d) in *Chemistry of Materials* and by Zoldesi and Imhof³ (Fig. 1e) in *Advanced Materials*. Their structures were fabricated by depositing silica on liquid crystals formed by surfactants², and through the regular deformation by osmotic stresses of thin siloxane shells grown around emulsion droplets that are monodisperse (all the same shape and size)³.

It is important to mention at this point that

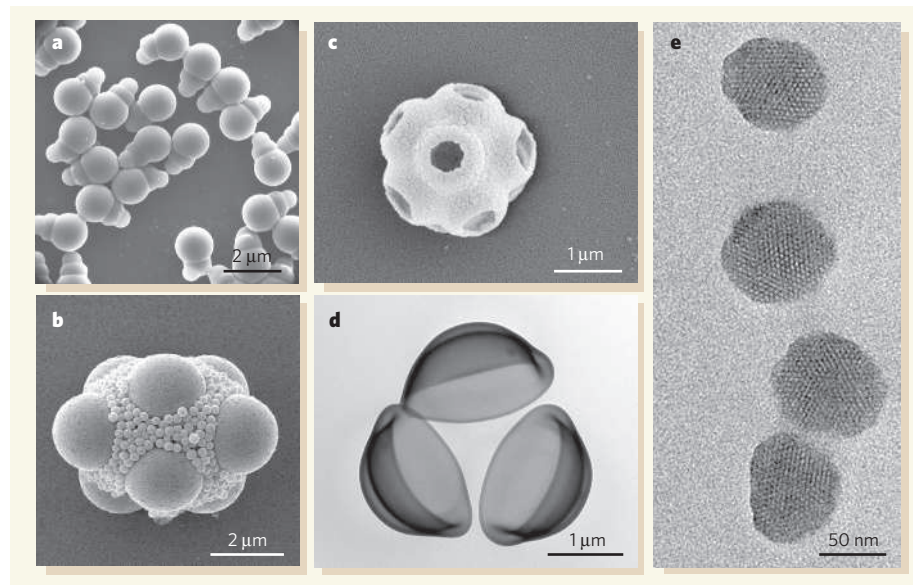


Figure 1 | A selection of complex colloids achieved by various means of self-organization. **a**, 'Ice-cream cones' resulting from repeated polymerization and phase separation between polymers of different composition⁴. **b**, **c**, Through controlled drying of a binary dispersion in water-in-oil emulsion droplets¹; **b**, both colloids same charge; **c**, colloids with opposite charge. **d**, Through silica deposition on liquid crystal phases formed by surfactants². **e**, Through osmotic stress deformation of thin hybrid siloxane shells after growing them on monodisperse oil droplets³. (Courtesy of: **a**, John Wiley, Inc.; **b**, **c**, American Chemical Society; **d**, C. M. van Kats, D. C. 't Hart and J. D. Meeldijk; **e**, C. I. Zoldesi and A. Imhof. All scale bars are approximate.)



50 YEARS AGO

"The training of university teachers" — The question of the advisability and possibility of providing new recruits to university teaching with some initial guidance in the technique of their calling has been examined by S. Radcliffe, lecturer in German at the University of Bristol... In general, lecturers are conscientious about the matter of their lectures, but give little thought to their form or their delivery... [Radcliffe] suggests that an artist requires some basic instruction, at least in the rudiments of his craft. The following are a few of the purely mechanical skills which might be considered desirable in a good teacher or lecturer. First, the adoption of a fitting speed and clarity of diction. Secondly, the clear formulation and appropriate stressing of the main points of the subject under review. Thirdly, the ability to use a blackboard successfully. Fourthly, the 'staging' of material to make it come 'alive'... Learning the students' names is an essential requirement in establishing closer contact with them... The prompt return of written work not only helps to keep up students' interest in their subject, but also gives the right to demand written work from the students within the time-limit specified.

From *Nature* 4 February 1956.

100 YEARS AGO

"The Revolution of the Corpuscle" A corpuscle once did oscillate so quickly to and fro, He always raised disturbances wherever he did go. He struggled hard for freedom against a powerful foe — An atom — who would not let him go. The aether trembled at his agitations In a manner so familiar that I only need to say, In accordance with Clerk Maxwell's six equations It tickled people's optics far away. You can feel the way it's done, You may trace them as they run — $d\gamma$ by dy less dz is equal $K.dX/dt$

From *Nature* 1 February 1906.

50 & 100 YEARS AGO

complex shape is no prerequisite for complex interaction. The adsorption of charged molecules on colloids with sizes of the order of micrometres⁵ and several nanometres⁶ has been shown, for example, to result in charges small enough that ionic colloidal crystals — crystals comprising particles of opposite charge — can form. The added complexity of the interactions between oppositely charged spheres, compared with components with the same charge, has already resulted in the number of different types of binary colloidal crystal that have been fabricated doubling within a few months^{5,6}.

Starting self-assembly with complex colloids, however, offers increased possibilities. This is nicely demonstrated by a modest goal that many groups are aiming for: the creation of colloidal crystals with diamond symmetry. Such structures could be used to create a photonic crystal with a robust 'band gap' that can inhibit the propagation of light and modify the spontaneous emission of photons at visible wavelengths. These crystals are potentially as useful for manipulating the flow of light waves as semiconductor crystals have become for manipulating the flow of electrons. The interest in these three-dimensional structures is so great that they have even been assembled do-it-yourself style by placing thousands of colloids into a diamond lattice one by one⁷. But, whereas just a few years ago most considered the creation of diamond lattices by self-assembly to be wishful thinking, several approaches now seem to make it an immediate prospect. Some proposed schemes, based on theory and computer simulations, make use of complex spherical colloids with either tetrahedrally arranged attractive patches⁸ or, remarkably, non-additive spherically symmetric potentials that might be produced using particles coated with complementary strands of DNA (that is, that can bind together to form double-stranded DNA)⁹.

For a second self-organization step to succeed in any system, all particles must be monodisperse and the yield of the first step must be high. The polydisperse emulsion droplets created by Cho *et al.*¹ are therefore at present unsuited for self-organization into more complex three-dimensional structures, because no two colloid particles that they produce are exactly the same. But the path to monodispersity, clearly outlined by the authors, should not be too arduous.

Going even further than this, a crystalline arrangement of smaller monodisperse nanoparticles between larger spheres, or a mixture of oppositely charged and monodisperse nanocrystals of different composition, could be created. Transistors with a conventional two-dimensional layout can be made from self-organized colloidal crystals of nanoparticles¹⁰; by self-organizing such semiconducting colloidal crystals between larger colloids, such as those devised by Cho *et al.*, individual transistors could be arranged into regular, three-dimensional structures. A three-dimensional

wiring system could also conceivably be established by using spheres that comprise a conducting and an insulating part as the larger building blocks, allowing each transistor to be addressed individually as it sits in the lattice.

Even though the complex colloidal structures now being created are beginning to show a faint resemblance to the beautiful silica structures produced by diatoms, natural examples of self-organization on multiple length scales and from different materials are generally still far ahead of any human design. Such structures are always available in case we run out of inspiration¹¹.

Alfons van Blaaderen is at the Debye Institute,

Utrecht University, Utrecht, The Netherlands.
e-mail: A.vanBlaaderen@phys.uu.nl

1. Cho, Y. S. *et al.* *J. Am. Chem. Soc.* **127**, 15968–15975 (2005).
2. Lin, Y. S. *et al.* *Chem. Mater.* **17**, 4570–4573 (2005).
3. Zoldesi, C. I. & Imhof, A. *Adv. Mater.* **17**, 924–928 (2005).
4. Sheu, H. R., El-Aasser, M. S. & Vanderhoff, J. W. *J. Polym. Sci. A* **28**, 629–651 (1990).
5. Leunissen, M. E. *et al.* *Nature* **438**, 235–239 (2005).
6. Shevchenko, E. V., Talapin, D. V., Kotov, N. A., O'Brien, S. & Murray, C. B. *Nature* **439**, 55–59 (2006).
7. Garcia-Santamaria, F. *et al.* *Adv. Mater.* **14**, 1144–1147 (2002).
8. Zhang, Z., Keys, A. S., Chen, T. & Glotzer, S. C. *Langmuir* **21**, 11547–11551 (2005).
9. Tkachenko, A. V. *Phys. Rev. Lett.* **89**, 148303 (2002).
10. Talapin, D. V. & Murray, C. B. *Science* **310**, 86–89 (2005).
11. Mann, S. *Biomimetic Materials: Principles and Concepts in Bioinorganic Materials Chemistry* (Oxford Univ. Press, 2001).

NEUROBIOLOGY

Memories of a fruitfly

William G. Quinn

Despite its tiny size, the fruitfly brain is staggeringly intricate. So teasing apart how it remembers things — even a simple line pattern — is a daunting task. Progress is being made, thanks to genetic innovations.

Neuroscientists these days have a satisfactory understanding of how individual neurons work and of how they communicate with their immediate neighbours. By contrast, understanding at the next level of organization is hazier; for example, how neurons form functional circuits, how these circuits encode behaviour and particularly how experience changes the activity and connectivity in circuits to alter behaviour.

In this fog, a natural question is: how simple a system can one study profitably? Molluscs such as the marine snail *Aplysia* have yielded much insight because they have large, simple neuronal circuits. However, these animals perform only very basic behaviours. Insects, on the other hand, often have intricate neural circuits and complex stereotyped behaviours, such as the dance language of the honeybee. But their neurons have seemed too small and tangled for conventional analyses. Advances in genetic techniques have overcome this problem of scale, and in this issue Liu *et al.* (page 551)¹ use these ingenious methods to begin to dissect finely how the fruitfly *Drosophila* learns visual patterns.

The authors take advantage of 'jumping genes', which can hop about the fly genome and splice themselves into chromosomes at random points. A jumping gene can be tailored to carry along another gene of interest — a transgene. If the transgene (called 'TG-1', say) is jumped into the DNA near a naturally occurring gene, it is usually expressed in the same tissues as the natural gene². This jumping-gene method is now highly developed in *Drosophila*, so that large numbers of fly

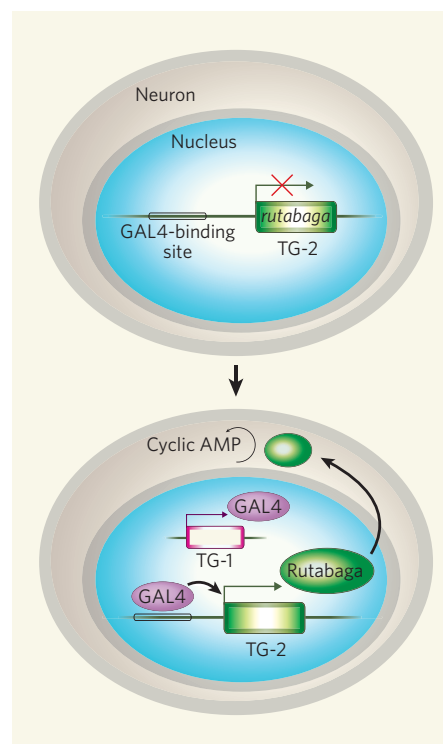


Figure 1 | Expressing a learning enzyme in selected cells. Liu *et al.*¹ used a strain of fruitfly that lacks a functional *rutabaga* gene, and added a substitute *rutabaga* gene (TG-2) that could be controlled by the transcriptional activator GAL4. By breeding these flies with various strains that express GAL4 only in certain subsets of neurons (from TG-1), they could ensure that the Rutabaga protein was produced only in those neurons. They then tested the flies' memories of various visual patterns.

FLUID DYNAMICS

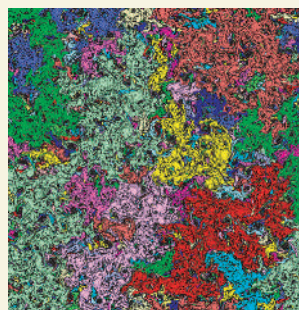
Flows like smoke and honey

Observe wafting cigarette smoke, and you are watching some complex physics: close to the lit end, the smoke rises smoothly; farther away, it unfurls into swirls. This second type of behaviour is an example of chaotic, turbulent flow, the course of which — because of the huge number of interacting particles involved — cannot be precisely predicted.

Denis Bernard *et al.* instead perform numerical simulations of turbulent flow (*Nature Phys.* **2**, 124–128; 2006). Surprisingly, they find a connection between two-dimensional turbulence and a simpler, hitherto unrelated

phenomenon — that of critical percolation.

Percolation describes the flow of a fluid through a porous material, such as honey seeping through beeswax. In a honeycomb, percolation reaches a critical threshold at which half the cells are filled with honey and the probability that a row of honey-filled cells spans the whole layer is not zero (although it might be infinitesimally small). At criticality, the filled and empty clusters of cells assume a 'fractal' pattern that follows a principle known as conformal invariance: as long as the angles don't change,



magnifying different sections of the pattern by different amounts results in a pattern that is indistinguishable from the original.

The principle is demonstrated in this image of Bernard and colleagues' turbulent flow. Here, connected clusters of vortices rotating in one direction are coloured and those rotating in

the other are black; the resulting pattern is indistinguishable on small and large scales. The authors calculate a fractal 'dimension', a measure of the degree to which the pattern seems to fill space as one looks at ever finer scales. The value they find, $4/3$, corresponds exactly to that found in the distribution of a fluid such as honey at the critical percolation threshold.

The discovery of this link could open the field of turbulence to the full theoretical artillery of conformal mapping, a technique that is central to a whole range of physical theories, such as string theory, besides critical percolation. But more work is required to smoke out the exact extent to which these ideas can be applied.

May Chiao

stocks are available by mail, each with TG-1 expressed in a different tissue, neuroanatomical structure or subgroup of neurons.

Nowadays, TG-1 usually encodes GAL4, a gene regulator that can be used specifically to turn on a second transgene ('TG-2') only in the requisite tissues or cell types. TG-2 will be a particular gene variant or mutation of interest, and it will have genetics such that it can be transferred into or out of a fly's genome via simple genetic crosses³. So it is now possible, by mail order and simple fly breeding, to express almost any interesting gene in an infinite variety of tissues or cells (Fig. 1). TG-2 can encode a gene product that identifies the selected neurons by dye or fluorescence, or it can be a protein that kills them, blocks transmission from them or overstimulates them. If TG-2 makes a protein that fluoresces in response to markers of neuronal activity such as elevated calcium or membrane voltage changes, then one can watch the neurons in action^{4,5}.

Liu *et al.*¹ used this method to dissect minutely the circuits involved in learning visual patterns. In their behavioural assay⁶, the fly faces a choice between two visual figures (for example, upright and inverted T shapes) and is punished by mild heating when it turns to one of them. Not surprisingly, the fly ends up facing the other figure; more surprisingly, it remembers to face the second figure in a subsequent test without heat reinforcement⁷.

Learning in *Drosophila* depends substantially on the Rutabaga enzyme. This enzyme is a calcium-calmodulin-dependent adenylyl cyclase⁸ that could be a molecular site for the convergence of signals from the cued stimulus and the reinforcement. Different types of learning require Rutabaga in different areas of the brain. Odour-discrimination learning, for instance, requires the enzyme found in brain structures called the mushroom bodies⁹. What Liu *et al.* show is that to learn specific visual

patterns, a fly must have Rutabaga in a brain region known as the central complex. They used mutant flies that lacked Rutabaga, and then switched Rutabaga on in a selection of specific patterns using the system above, and tested to see which flies remembered the visual cues. Strikingly, discrimination between different features — the angular orientation of a bar or its vertical position on the fly's retinal field — requires Rutabaga in different sets of central-complex neurons, which project to different layers of the 'fan-shaped body' (FB), a substructure of the central complex (Fig. 2). Those neurons that use Rutabaga to store vertical elevation extend into FB layer 5; those that encode contour orientation project to FB layer 1.

How does this work advance the field? Memory has been plausibly localized to small groups of neurons before, in *Aplysia*, but that was memory simply to increase or decrease

the strength of an innate reflex — gill withdrawal after electric shock. Liu *et al.* localize memories much richer in information about higher-order visual properties — memories that presumably require the cells' integration into finely tuned circuits. Can we understand the circuits? Probably not for a while. The neuroanatomy of the central complex has been systematically studied, and several stereotyped classes of neuron have been categorized and mapped there¹⁰. The organization of the central complex is repetitive and regular, but fiendishly intricate. Understanding how the circuitry encodes angular orientation, for example, will require either new methods or clever, and lucky, anatomical insight.

A more tractable question is: how does Rutabaga bring about changes in the appropriate central-complex cells to encode the memories, and what are the relevant changes? If the changes are structural, there are methods that will pinpoint them¹¹. Finding electrical changes in the cells is a knottier problem, because the central-complex cells are small and located deep inside the fly brain. Clues may come from other mutants deficient in learning, work on *Aplysia*, or studies of larger, more superficial cells in the *Drosophila* brain⁵. What Liu *et al.* make very clear is that genetic trickery has converted *Drosophila* from one of the worst organisms for functional neuroanatomy to one of the better ones. Indeed, for localizing individual, complex memories it may be the best. ■

William G. Quinn is in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. e-mail: cquinn@mit.edu



Figure 2 | Pinpointing visual memories. The pictures show brain sections of flies, with the fan-shaped body (FB) outlined in yellow. In the cells coloured white, the Rutabaga enzyme was needed to learn the visual cues shown on the left. Liu *et al.*¹ show that the ability to learn the horizontal elevation of a cue requires Rutabaga in a different set of FB neurons than does the ability to learn a diagonal cue. The involvement of the white-stained cells seen elsewhere in the brain was ruled out by other experiments. (Adapted from Liu *et al.*¹.)

1. Liu, G. *et al.* *Nature* **439**, 551–556 (2006).
2. O'Kane, C. J. & Gehring, W. *Proc. Natl Acad. Sci. USA* **84**, 9123–9127 (1987).
3. Brand, A. H. & Perrimon, N. *Development* **118**, 401–415 (1993).

4. Suh, G. S. *et al.* *Nature* **431**, 854–859 (2004).
5. Yu, D. *et al.* *Cell* **123**, 545–557 (2005).
6. Heisenberg, M. & Wolf, R. J. *Comp. Physiol. A* **130**, 113–130 (1979).
7. Dill, M. *et al.* *Nature* **365**, 751–753 (1993).
8. Sziber, P. P. *et al.* *Cell* **37**, 205–215 (1984).
9. Zars, T. *et al.* *Science* **288**, 672–675 (2000).
10. Hanesch, U., Fischbach, K. F. & Heisenberg, M. *Cell Tissue Res.* **257**, 343–366 (1989).
11. Lee, T. & Luo, L. *Trends Neurosci.* **24**, 251–254 (2001).

CHEMISTRY

Catalysts live and up close

Bert M. Weckhuysen

Designing efficient solid-state catalysts would be easier if we knew which parts of them do what. Fluorescence microscopy could help: the technique allows single catalytic events to be observed in real time.

Catalysts are the workhorses of the chemical industry: more than 80% of all modern chemicals come into contact with at least one catalyst during their manufacture. They can also be extremely complex: solids with large surface areas, for example, possess many potential active sites in their crystal structure. In this issue, Roefsaers *et al.*¹ (page 572) premiere a method of 'filming' single catalytic events in real time on a solid crystal, allowing catalytic activity to be mapped over its whole surface. The innovation is a step further on the road to the 'rational design' of catalysts, which offers the prospect of improved formulations for existing catalysts, and more effective and selective catalysts created from scratch.

Rational design remains in most cases a pipe-dream: the experimental tools available for monitoring catalysts in action are still, in the main, too rudimentary. Nevertheless, this area of research — often referred to as *in situ* spectroscopy — has seen tremendous progress over the past decade, partially as a result of improvements in analytical instrumentation^{2–6}. The latest contribution¹ exploits a method known as fluorescence microscopy. This technique has been used, for instance, in combinatorial catalysis, to screen whole series of catalyst libraries for the formation of fluorescence, and for monitoring the disappearance of specific organic molecules during catalytic transformations^{7–9}. It has also been used to track the diffusion patterns of fluorescent molecules, and to determine their diffusion coefficients in different porous oxides^{10,11}. Furthermore, fluorescence microscopy can monitor the adsorption and desorption of a dye molecule by individual crystals of a catalyst material¹².

The distinctive aspect of Roefsaers and colleagues' work¹, however, is the use of high-resolution fluorescence microscopy to observe a catalytic crystal at work in liquid-phase reactions. The high sensitivity of their approach enabled the authors to monitor single catalytic events in real time by observing a 'reporter' molecule that becomes fluorescent only after catalytic action. They could thus beautifully map the spatial distribution of active sites over a single catalytic crystal.

The authors tested the method on a layered double hydroxide (LDH) catalyst consisting of prismatic crystals with large basal planes and lateral faces (see Fig. 1 on page 573). Intriguingly, the tracks of the fluorescing molecules revealed that one reaction, ester hydrolysis, is catalysed only by active sites on the lateral faces of the LDH particle, whereas another reaction, transesterification, occurs on the entire outer crystal surface. In the terminology of surface scientists, ester hydrolysis is a structure-sensitive reaction, whereas transesterification is not.

Such 'crystal-face-dependent' catalysis has already been observed on metal single-crystal surfaces, for example in the synthesis of ammonia, where the rate of product formation is found to depend critically on the orientation of the crystalline iron catalyst¹³. But the current work is, I believe, the first of its kind to use catalytic solids in liquid-phase applications.

Whether fluorescence microscopy can be introduced as a more general method for *in situ* spectroscopy in catalysis research clearly depends on the availability of a sufficient range of fluorescent molecules that respond to different reactions. Attention should also be paid to the question of whether fluorescent reporter molecules behave like non-fluorescent molecules, and so indeed reproduce the distribution of catalytic sites reliably. Finally, fluorescence quenching — which may significantly affect the intensity of fluorescent emission — could become a problem, especially when the approach is extended to more severe reaction conditions.

Currently, work on the *in situ* spectroscopy of catalytic solids is divided, roughly speaking, into two groups, probably reflecting the differing expertise of the chemists involved. The first group focuses on the inorganic part of the catalyst material, and aims to capture its oxidation state and the geometrical structure of its bonds (its 'coordination environment'). This group uses techniques such as absorption spectroscopy and electron paramagnetic resonance. The second group, which focuses on the organic parts of the catalyst, uses nuclear magnetic resonance and vibrational

spectroscopic techniques, such as infrared and Raman spectroscopy, to illuminate reaction mechanisms and potential reaction intermediates.

With Roefsaers and colleagues' work¹, the organic chemists gain a new tool for their *in situ* spectroscopic work. Further efforts should be directed towards combining both schools of thought, focusing, with nanometre resolution, on both the inorganic and organic parts of a catalyst. That could allow crucial insights into changes in the coordination environment of an active site, as well the reaction intermediates formed. The results could then be interpreted in terms of catalytic reactivity and selectivity in a 'joined-up' fashion. Although combining two or more spectroscopic techniques might seem simple, many hurdles must be cleared before it becomes a reality^{14,15}.

But the main challenge is to push the resolution of *in situ* spectroscopic tools into the nanometre range. Transmission electron microscopy offers resolution at the atomic scale (around a tenth of a nanometre). Coupling this with electron energy-loss spectroscopy, in which the sample is bombarded with a monoenergetic beam of electrons, allows chemical mapping of a catalytic surface. However, the energy resolution of this technique is still too low to identify the different oxidation states of the bombarded material. Moreover, electron-based techniques also need low pressures (in the millibar range), and thus the results obtained may deviate from those of experiments performed under true reaction pressures.

So there is room for other spectroscopic tools that, in combination with Roefsaers and colleagues' fluorescence microscopy¹, could deliver both inorganic and organic information about a catalytic process, creating a powerful *in situ* nanospectroscopy technique for taking snapshots of a wide variety of catalytic solids at work. If this endeavour succeeds, rational catalyst design will be a step closer to realization. ■

Bert M. Weckhuysen is in the Department of Chemistry, Debye Institute, Utrecht University, Sorbonnelaan 16, PO Box 80083, 3508 TB Utrecht, the Netherlands.
e-mail: b.m.weckhuysen@chem.uu.nl

1. Roefsaers, M. B. J. *et al.* *Nature* **439**, 572–575 (2006).
2. Haw, J. F. (ed.) *In-Situ Spectroscopy in Heterogeneous Catalysis* (Wiley-VCH, Weinheim, 2002).
3. Weckhuysen, B. M. (ed.) *In-Situ Spectroscopy of Catalysts* (Am. Sci., Stevenson Ranch, CA, 2004).
4. Weckhuysen, B. M. *Chem. Commun.* 97–110 (2002).
5. Hunger, M. & Weitkamp, J. *Angew. Chem. Int. Edn* **40**, 2954–2971 (2001).
6. Thomas, J. M. *Angew. Chem. Int. Edn* **38**, 3589–3628 (1999).
7. Shaughnessy, K. H., Kim, P. & Hartwig, J. F. *J. Am. Chem. Soc.* **121**, 2123–2132 (1999).
8. Copeland, G. T. & Miller, S. J. *J. Am. Chem. Soc.* **121**, 4306–4307 (1999).
9. Su, H. *et al.* *Anal. Chem.* **73**, 4434–4440 (2001).
10. Hellriegel, C. *et al.* *N. J. Phys.* **7**, 23 (2005).
11. Seebacher, C. *et al.* *J. Phys. Chem. B* **106**, 5591–5595 (2001).
12. Roefsaers, M. B. J. *et al.* *ChemPhysChem* **6**, 2295–2299 (2005).
13. Spencer, N. D. *et al.* *J. Catal.* **74**, 129–135 (1982).
14. Bruckner, A. *Chem. Commun.* 1761–1763 (2005).
15. Beale, A. M. *et al.* *Chem. Commun.* 3015–3017 (2005).

BRIEF COMMUNICATIONS

Transmission of devil facial-tumour disease

An uncanny similarity in the karyotype of these malignant tumours means that they could be infective.

The Tasmanian devil, a large carnivorous Australian marsupial, is under threat from a widespread fatal disease in which a malignant oral-facial tumour obstructs the animal's ability to feed¹. Here we show that the chromosomes in these tumours have undergone a complex rearrangement that is identical for every animal studied. In light of this remarkable finding and of the known fighting behaviour of the devils², we propose that the disease is transmitted by allograft, whereby an infectious cell line is passed directly between the animals through bites they inflict on one another.

The cancer, known as devil facial-tumour disease, now affects devils (*Sarcophilus harrisii*) in more than half of Tasmania¹. The growth of the tumours, which ulcerate and become friable, eventually causes the devils to starve. As the tumour cells are easily dislodged and because almost all bites from the devils' frequent fighting occur around the mouth², we investigated whether the disease might be transmitted by allograft between animals. We studied tumours that included early neoplasms, huge primary cancers and secondary cancers. The cancers were sampled from animals throughout eastern Tasmania, Australia, over a 12-month period (for methods, see supplementary information).

The normal number of chromosomes in the devil is 14, including the XX or XY sex chromosomes (Fig. 1a). We found that the facial tumours contained only 13 chromosomes and that these were grossly abnormal (Fig. 1b). The number and appearance of the chromosomes (the karyotype) indicated that both sex chromosomes, both chromosomes 2 and one chromosome 6 were absent. There was also a deletion of the long arm of one chromosome 1, and four unidentified marker chromosomes were present. Most important, these anomalies were the same in the facial tumours from every animal ($n = 11$).

These rearrangements are complex, but no intermediate stages were found between normal and tumour chromosomes, even in small primary cancers. In human cancers, there is generally a common breakpoint (first event)³, irrespective of whether the neoplasm is caused by viral insertion (as in Burkitt's lymphoma⁴) or arises spontaneously (as in Ewing's sarcoma⁵); complex rearrangements occur in solid tumours as a result of further clonal evolution⁵. However, the identical chromosomal rearrangements that we found in the facial

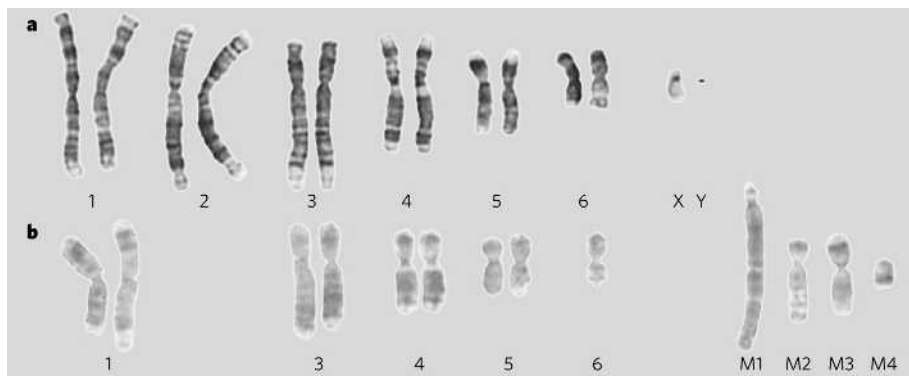


Figure 1 | Chromosomes of facial tumours from Tasmanian devils. **a**, Normal karyotype for a male Tasmanian devil (14 chromosomes, including XY). **b**, Karyotype of cancer cells found in each of the facial tumours of all 11 animals studied (13 chromosomes, with no sex chromosomes, no chromosome-2 pair and only one chromosome 6; the long arm of one chromosome 1 was deleted; four additional marker chromosomes were present (M1–M4)).

tumours of each devil are too complex for a common breakpoint to have occurred³. Indeed, the rearrangements do not conform to any human model, particularly given the loss of sex chromosomes in all the tumours of devils of both sexes.

Further support for the allograft theory of disease transmission derives from the serendipitous observation of a pericentric inversion of chromosome 5 in the constitutional karyotype of one animal. This constitutional anomaly was found in all cultures of that devil's normal tissues, but was not present in either of the chromosomes 5 in his facial-tumour cells, where it would have been found had the neoplasm arisen from his own tissue.

Cases of transmissible venereal sarcoma in dogs⁶ also show similar chromosomal defects among tumours, leading to the proposition that this sarcoma may develop from "a clone capable of a parasitic existence"⁷ — a description that also fits the features of the devil's facial tumour. We suggest that the devils' cancer (like the dogs') is infective and that the infective agent is a rogue cell line that initially evolved in a tumour of unknown origin.

Humans, too, can accidentally infect each other with cancer, through cell implantation in patients that have received organ transplants⁸; such cancers then develop according to their usual course⁹. Organ transplants are less likely to be rejected if the donor is a close relative who has a matching tissue type; by analogy, the low genetic diversity and high degree of kinship among devils¹⁰ might help to reduce their immune response to cancer cells implanted

during biting. Although the devil's immune system is poorly understood, preliminary investigations indicate that there is little immune reaction between lymphocytes taken from devils from within and outside local populations (G. Woods, personal communication).

To obtain further insight into the transmission of the devil's facial-tumour disease, it will be necessary to DNA-fingerprint tumours and clarify their derivation by using whole-chromosome painting probes, as well as searching for oncogenes. This should reveal the disease's toxicology, progression and epidemiology.

A.-M. Pearce, K. Swift

Department of Primary Industries, Water and Environment, PO Box 46, Kings Meadows, Tasmania 7249, Australia
e-mail: anne-maree.pearce@dpiwe.tas.gov.au

1. Hawkins, C. E. et al. *Biol. Cons.* (submitted).
2. Hamede, R. Thesis, Univ. Tasmania (Hobart, 2004).
3. Mitelman, F. *Mutat. Res.* **462**, 247–253 (2000).
4. Zech, L., Haglund, U., Nilsson, K. & Klein, G. *Int. J. Cancer* **17**, 47–56 (1976).
5. Mitelman, F., Johansson, B., Mandahl, N. & Mertens, F. *Cancer Genet. Cytogenet.* **95**, 1–8 (1997).
6. Das, U. & Das, A. K. *Vet. Res. Commun.* **24**, 545–556 (2000).
7. Murray, M., James, Z. H. & Martin, W. B. *Res. Vet. Sci.* **10**, 565–568 (1969).
8. Kauffman, H. M. & McBride, M. A. *Transplantation* **74**, 358–362 (2002).
9. Forbes, G. B. et al. *J. Clin. Pathol.* **34**, 109–115 (1981).
10. Jones, M. E., Paetkau, D., Geffen, E. & Moritz, C. *Mol. Ecol.* **13**, 2197–2209 (2004).

Supplementary information accompanies this communication on Nature's website.

Received 27 September 2005; accepted 17 January 2006.

Competing financial interests: declared none.

doi:10.1038/439549a

Distinct memory traces for two visual features in the *Drosophila* brain

Gang Liu^{1,2*}, Holger Seiler^{3*}, Ai Wen¹, Troy Zars^{3,4}, Kei Ito⁵, Reinhard Wolf³, Martin Heisenberg³ & Li Liu¹

The fly *Drosophila melanogaster* can discriminate and remember visual landmarks. It analyses selected parts of its visual environment according to a small number of pattern parameters such as size, colour or contour orientation, and stores particular parameter values. Like humans, flies recognize patterns independently of the retinal position during acquisition of the pattern (translation invariance). Here we show that the central-most part of the fly brain, the fan-shaped body, contains parts of a network mediating visual pattern recognition. We have identified short-term memory traces of two pattern parameters—elevation in the panorama and contour orientation. These can be localized to two groups of neurons extending branches as parallel, horizontal strata in the fan-shaped body. The central location of this memory store is well suited to mediate translational invariance.

Drosophila tethered to a torque meter, with its head (and hence its eyes) fixed in space, can control its orientation with respect to the artificial scenery in a flight simulator¹. In this set up, the fly is conditioned to avoid certain flight directions relative to virtual landmarks (Fig. 1) and recognizes these visual patterns for up to at least 48 h (ref. 2). Visual pattern recognition in *Drosophila* has been studied in some detail^{3–7}. Flies store values of at least five pattern parameters: size, colour, elevation in the panorama, vertical compactness, and contour orientation. Moreover, they memorize spatial relations between parameter values. The neuronal substrate underlying visual pattern recognition is little understood in any organism.

In *Drosophila*, memory traces can be localized to groups of neurons in the brain⁸. Using the enhancer GAL4/UAS expression system^{9,10}, short-term memory traces of aversive and appetitive olfactory conditioning have been assigned to output synapses of subsets of intrinsic neurons of the mushroom bodies (MBs). The Rutabaga protein—a type 1 adenylyl cyclase that is regulated by Ca²⁺/Calmodulin and G protein, and is considered a putative convergence site of the unconditioned and conditioned stimulus in olfactory associative learning^{11–14}—selectively restores olfactory learning if expressed in these cells in an otherwise *rutabaga* (*rut*)-mutant animal^{15,16}. Moreover, expressing a mutated constitutively activating G α_s protein (G α_s^*) in the MBs interferes with olfactory learning¹⁷. Blocking the output from these neurons during memory retrieval has the same effect, while blocking it during acquisition has no effect^{16,18,19}. Interestingly, memory traces for other learning tasks seem to reside in other parts of the brain: for remembering its location in a dark space, the fly seems to rely on a *rut*-dependent memory trace in neurons of the median bundle and/or the ventral ganglion²⁰.

In the present study, we localize short-term memory traces for visual pattern recognition to the fan-shaped body (FB), the largest component of the central complex (CX; also called the central body in other species). The CX is a hallmark of the arthropod brain. It has been characterized functionally as a pre-motor centre with prominent, but not exclusive, visual input (for a review see refs 21, 22). In

the locust, large-field neurons sensitive to the *e*-vector orientation of polarized light have been described in the CX²³. Because of its repetitive structure and the precisely ordered overlay of fibre projections from the two hemispheres in the FB, neighbourhood relations of visual space might still be partially preserved at this level (retinotopy)²⁴. Using the genetic approach^{15,16,20}, we now show that a small group of characteristic stratified neurons in the FB house a memory trace for the pattern parameter ‘elevation’, and a different set of neurons forming a parallel stratum contain a memory trace for ‘contour orientation’.

Central complex defects impair visual pattern memory

Of ten mutants with structural abnormalities in the CX, all were impaired in visual pattern recognition^{25,26}. They were able to fly straight and to avoid heat, yet they failed to remember the patterns (Fig. 2a). Did they really lack the memory or had they lost their ability to discriminate between patterns? Fortunately, individual flies often display spontaneous preferences for one of the patterns (see Fig. 2c and the Supplementary Information for the evaluation procedure (Table S1)). In three lines, these preferences were consistent enough to reveal intact pattern discrimination, suggesting that aberrant circuitry of the central complex can affect visual learning independent of visual pattern discrimination.

As the developmental and structural defects in these mutants are not well characterized, we used the GAL4/UAS system to acutely interfere with CX function. We chose a GAL4 driver line (c205–GAL4) with expression in parts of the CX (expression pattern shown in Fig. 3c) and, as the effector, the gene for tetanus toxin light chain (CnTE)²⁷. CnTE blocks neurons by cleaving neuronal Synaptobrevin, a protein controlling transmitter release²⁸. For temporal control, we added the temperature-sensitive GAL4-specific silencer GAL80 under the control of a *tubulin* promoter (*tub*–GAL80^{ts})¹⁰. Flies (UAS–CnTE/+; *tub*–GAL80^{ts}/c205–GAL4) were raised at 19 °C, and were transferred for 14 h to the restrictive temperature (30 °C) just before the behavioural experiment to induce GAL4-driven toxin expression. Flies kept at the low temperature showed normal

¹State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China. ²Graduate School of the Chinese Academy of Sciences, Beijing 100039, China. ³Theodor-Boveri-Institut für Biowissenschaften, Lehrstuhl für Genetik und Neurobiologie, Am Hubland, 97074 Würzburg, Germany. ⁴Division of Biological Sciences, 219 Lefevre Hall, University of Missouri, Columbia, Missouri 65211, USA. ⁵Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.

*These authors contributed equally to this work.

memory scores, while after inactivation of $GAL80^{ts}$ no pattern memory was observed (Fig. 2b). Again, flight control and heat avoidance were normal, and Fourier analysis confirmed that flies at the high temperature had retained their ability to tell the patterns apart (Supplementary Table S1). As with the structural mutants, interrupting the circuitry of the CX by tetanus toxin expression seemed to specifically interfere with visual pattern memory. The use of *tub-GAL80^{ts}*, in addition, excluded the possibility that toxin expression in unknown tissues during development might cause the memory impairment in the adult. These results do not, as yet, address the question of memory localization.

The *rutabaga* gene is required for pattern memory

Visual pattern memory in the flight simulator requires an intact *rut* gene (Fig. 2c, left)²⁹. Mutant *rut* flies (*rut²⁰⁸⁰*) showed normal visual flight control, heat avoidance and pattern discrimination (Fig. 2d). To confirm that the defect was indeed due to the mutation in the *rut* gene rather than an unidentified second-site mutation, we rescued it by the expression of the wild-type *rut* cDNA (*UAS-rut⁺*) using the

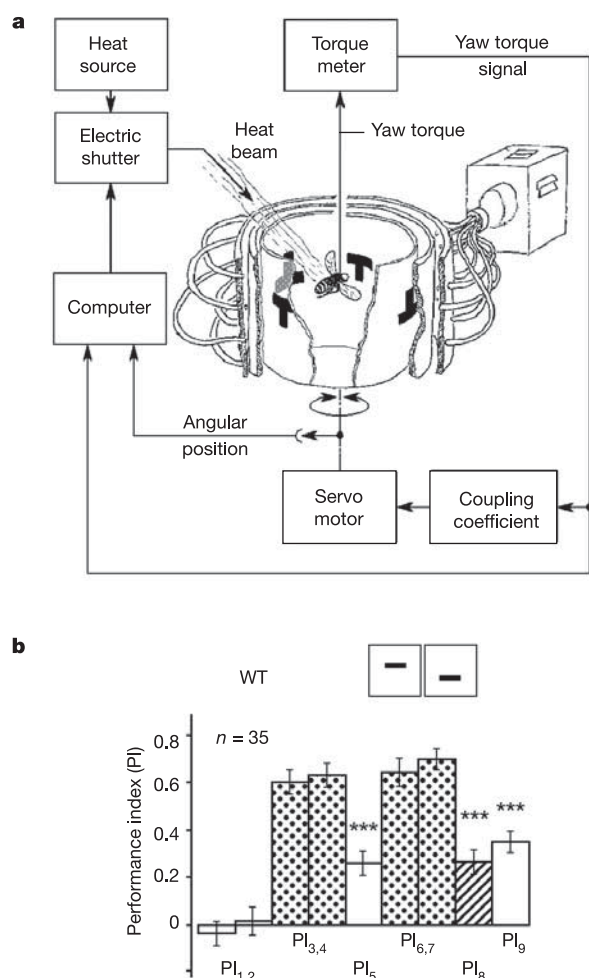


Figure 1 | Flight simulator for measuring visual pattern recognition.

a, Experimental setup. The fly is attached to a torque meter and its yaw torque controls the angular velocity of the panorama surrounding it. It can fly straight and chooses its flight direction relative to the visual patterns in the panorama. With a beam of infrared light, the fly can be conditioned to avoid certain flight directions (see the Methods for more details). **b**, Course of experiment. Bars show performance indices (PI) of successive 2-min intervals of pretest (PI_1, PI_2), training (dotted bars; PI_3, PI_4, PI_6, PI_7) and memory test (PI_5, PI_8, PI_9) (see the Methods for experimental details and definition of PI). The following figures all show PI_8 (hatched bar) exclusively. Error bars are s.e.m. throughout. WT, wild-type *Berlin* strain.

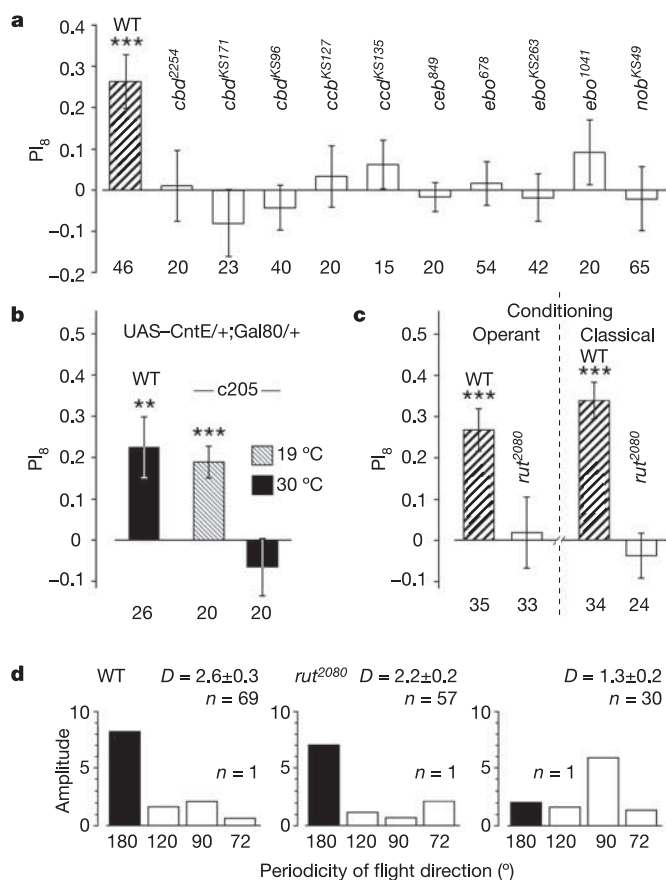


Figure 2 | Visual pattern memory is impaired in central complex mutants.

Memory (PI_8) for the pattern parameter 'elevation' was tested using either horizontal bars at different elevations in the panorama or upright and inverted Ts. (The fly can conditionally distinguish the latter only by the different elevations of their centres of gravity⁶.) **a**, Visual pattern memory in wild-type (WT) *Berlin* flies (hatched bar) and ten CX mutants in six independent genes. Mutants are in the genetic background of WT *Berlin*. Experimental details can be found in the Methods, Supplementary Information and Fig. 1. None of the mutant memory scores are significantly different from zero. **b**, Expression of tetanus toxin light chain (CnTE) in the driver line *c205-GAL4* labelling parts of the CX interferes with visual pattern memory. Owing to the *tub-GAL80^{ts}* element, a conditional silencer of *GAL4*, CnTE is expressed only in the adult during a 14 h incubation at 30 °C (the restrictive temperature for *tub-GAL80^{ts}*) before the experiment. *GAL4* lines are derived from wild-type Canton-S. As WT control, therefore, Canton-S was chosen. **c**, Visual pattern memory is impaired in the mutant *rut²⁰⁸⁰*. Memory scores after operant (left) and classical (pavlovian) conditioning (right). Note that during classical conditioning, flies are not in command of panorama motion and are exposed to heat for precisely 50% of the time ($PI = 0$). Panorama motion alternates between smooth clockwise and anticlockwise rotation, each for 12 s at an angular velocity of 30°sec^{-1} . Memory test is the same as in the operant experiment. **d**, Pattern discrimination in the flight simulator. If flies show spontaneous individual pattern preferences, orientation histograms (not shown) have a 180° periodicity owing to identical patterns in opposing quadrants. Left and middle panels show principal Fourier coefficients during the 4-min pre-test period for single WT (left) and *rut* (middle) flies. The 180° Fourier components (black bars) have the largest amplitude. The same evaluation of a 4-min pre-test with four identical patterns (upright Ts) is also shown (right panel). The 180° component is at background level, and the high 90° component indicates equal choice for all four quadrants. Discrimination values (D) are calculated as $D = 2A_{180}/(A_{120} + A_{72})$ (subscripts refer to Fourier components; A = amplitude). Numbers for D above Fourier spectra are averages $\pm$ s.e.m. for n flies (for further details, see the Supplementary Information). Numbers below bars in **a–c** indicate numbers of flies. Error bars are s.e.m.

pan-neuronally expressing driver line *elav*-GAL4. Indeed, flies of the genotype *rut*²⁰⁸⁰/Y;*elav*-GAL4/UAS-*rut*⁺ had normal memory (Supplementary Table S2A).

Visual pattern memory in the flight simulator has been shown to depend upon at least two kinds of behavioural plasticity. For one, an associative classical (pavlovian) memory trace is formed linking a particular set of values of pattern parameters to heat. Second, the fly's control of the panorama operantly facilitates the formation of this memory trace³⁰. Either of the two processes might depend upon the Rut cycle.

To address this issue, we tested *rut* mutant flies in a purely classical variant of the learning paradigm. During training, panorama motion was uncoupled from the fly's yaw torque and the panorama was slowly rotated around the fly. Heat was made contingent with the appearance of the 'punished' pattern in the frontal quadrant of the fly's visual field. All other parameters were kept as described (see the Methods). For testing memory, panorama motion was coupled again to yaw torque and the fly's pattern preference was recorded as usual. Even in the absence of operant facilitation, visual pattern memory required the intact *rut* gene (Fig. 2c, right). Therefore, the *rut*-dependent memory trace investigated here represents the association of a property of a visual pattern with the reinforcer.

Local rescue of visual pattern memory in the *rut* mutant

As a first step in localizing the memory trace, we asked in which neurons of the *rut* mutant expression of the wild-type *rut* gene would be sufficient to restore learning. To this end, a total of 27 driver lines expressing GAL4 in different neuropil regions of the brain were used to drive the UAS-*rut*⁺ effector gene in the *rut* mutant background. The parameter 'elevation' was measured. With seven of the driver lines, pattern memory was restored (104y, 121y, 154y, 210y, c5, c205 and c271; Fig. 3 and Supplementary Fig. S2; Supplementary Table S2B, C).

Comparison of the expression patterns of the 27 lines allowed us to narrow down the putative site of the memory trace to a small group of neurons in the brain. The seven rescuing lines all showed transgene expression in a stratum in the upper part of the FB. In three of them staining is rather selective. It comprises, in addition to the FB, only a layer in the medulla, several cell clusters in the suboesophageal ganglion and a few other scattered neurons.

Evidently, *rut*⁺ expression in the MBs is neither necessary (104y, c5, c205, 154y) nor sufficient for rescue (Supplementary Table S2B). This result is in line with the earlier observation that elimination of more than 90% of the MBs by hydroxyurea treatment of first-instar larvae³¹ has no deleterious effect on visual pattern memory³². We ablated the MBs in one group of rescue flies (*rut*²⁰⁸⁰/Y;UAS-*rut*⁺/+;c271/+). They showed full visual pattern memory (data not shown).

Although GAL4 expression in the optic lobes is prominent in all seven rescuing lines, it occurs in distinctly different layers that do not overlap (Supplementary Fig. S2). For instance, in 104y expression is restricted to layer 2, whereas in 210y it is found only in the serpentine layer (layer 7)³³. A similar situation is found for the suboesophageal ganglion, although there the staining patterns are more difficult to evaluate. Finally, expression in the ellipsoid body is again not necessary (104y, c5, c205, 154y) or sufficient (c232, 78y, 7y, and so on) for rescue. Thus, the expression patterns favour the conclusion that the neurons of the upper stratum of the FB (Supplementary Fig. S2) might be the site of the memory trace for the parameter 'elevation' in visual pattern memory.

Neurons in this stratum, labelled in all seven rescuing lines, have a very characteristic shape (Supplementary Fig. S3). Their cell bodies are located just lateral to the calyces. Their neurites run slightly upward in an antero-medial direction, forming an upward-directed tufted arborization just behind the α/α' -lobe of the MB. From there, the fibre turns sharply down and backward towards the midline just in front of the FB. Finally, it turns horizontally backward, spreading as a sharp stratum through all of the FB across the midline. These

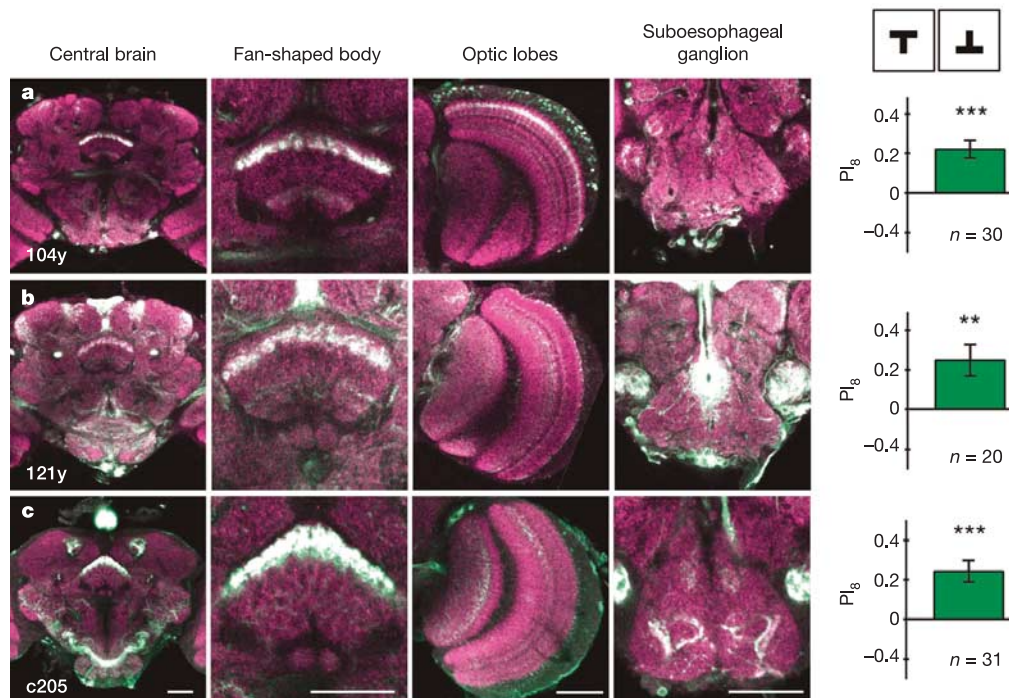


Figure 3 | Expression patterns of three driver lines rescuing pattern memory. GAL4-driven green fluorescent protein (GFP) expression in three of the seven 'rescuing' driver lines. Flies mutant for *rut* show normal visual pattern memory (PI₈, green bars on the right), (a), if 104y-GAL4, (b) 121y-GAL4 or (c) c205-GAL4 drive *rut*⁺ cDNA. Frontal virtual sections of

whole-mount preparations stained with anti-nc82 antibody (staining the synaptic neuropil, magenta) and anti-GFP antibody (GAL4 expression, green; overlay white). Scale bars 50 μ m. All 'rescuing' driver lines (see Supplementary Fig. S2) show labelling in a dorsal horizontal layer (layer 5) of the FB. Error bars are s.e.m.

neurons have been described before in Golgi preparations³⁴. They belong to a larger group of tangential FB neurons called Fneurons³⁴. Besides the stratum in FB, most of them have an arborization in a particular part of the unstructured neuropil. We tentatively classify the layer stained in 104y, and the other six rescuing lines, as layer 5 (from bottom upward), and hence provisionally call the neurons F5, although, without further markers, it is difficult to reliably number the layers. In summary, expression of Rut cyclase in F5 neurons rescued the *rut*-dependent memory defect for pattern elevation, whereas no rescue effect was observed in any of 20 strains without expression of Rut cyclase in F5 neurons (though Rut cyclase was expressed in other regions of the brain). Hence, a *rut*-dependent memory trace for pattern elevation may reside in F5 neurons.

This finding does not exclude the possibility that memory is redundant, and that other *rut*-dependent memory traces for pattern elevation might be found elsewhere. Therefore, we asked whether plasticity in the F5 neurons is necessary for visual pattern memory. As mentioned above, the Rut cyclase is regulated by G protein^{11–14}, and olfactory learning/memory can be blocked by a constitutively active form of the $G\alpha_s$ protein subunit ($G\alpha_s^*$)¹⁷. We expressed the $G\alpha_s^*$ mutant protein in the FB using the driver line c205, and tested the flies for their memory of 'elevation'. Memory was fully suppressed (Fig. 4a). As in olfactory learning¹⁷, overexpression of the wild-type protein did not interfere with learning (Fig. 4a). These results support the hypothesis that continuous upregulation of Rut cyclase in the F5-neurons interferes with visual short-term memory, implying that F5 neurons are the only site of a *rut*-dependent memory trace for pattern elevation.

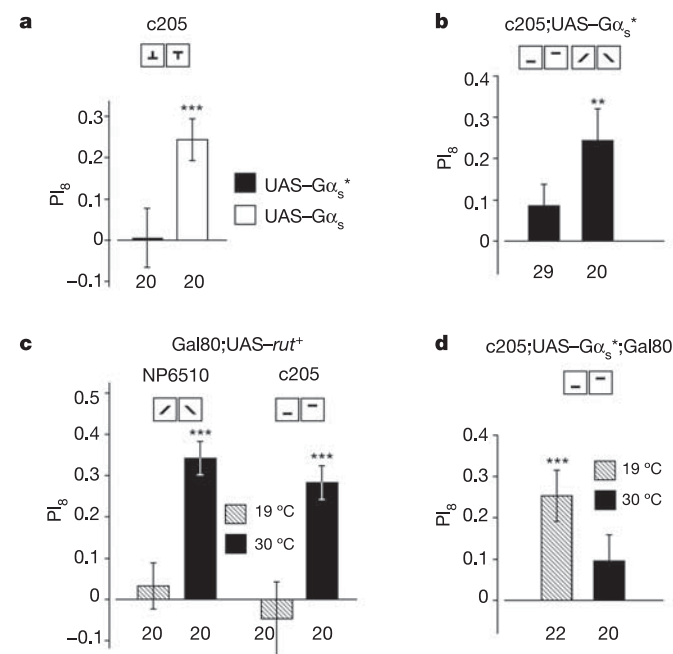


Figure 4 | Rescue and suppression of pattern memory occur in the adult.

a, Expression of a constitutively active $G\alpha_s^*$ protein subunit in the FB (driver line c205-GAL4) abolishes pattern memory. Overexpression of wild-type $G\alpha_s$ protein has no effect. **b**, $G\alpha_s^*$ driven by c205-GAL4 disrupts memory for 'elevation' but leaves memory for 'contour orientation' intact. **c**, Temporal specificity of pattern memory. Expression of *rut*⁺ cDNA in a mutant *rut* background is blocked during development using *tub*-GAL80^{ts} (light hatched bars). Only after inactivation of GAL80^{ts} at 30 °C in the adult, Rut expression is induced and pattern learning/memory restored (black bars). **d**, $G\alpha_s^*$ protein interferes with pattern memory if expression is delayed until adulthood. Temperature control as in panel **c**. Numbers below the bars indicate numbers of flies. Error bars are s.e.m.

Pattern specificity of the memory trace

The patterns used in the experiments so far exclusively addressed the parameter 'elevation' (upright and inverted Ts or horizontal bars at different elevations). We wondered whether the mutant defect in *rut* and the Rut rescue in the F5 neurons affected only this parameter, or whether it applied to other pattern parameters as well. We therefore extended the study to two further parameters: 'size' and 'contour orientation'. Three driver lines—c205, NP6510 and NP2320—were chosen showing different expression patterns in the FB (Fig. 5a–c). In the line NP6510, as in c205, a group of F neurons is marked. They are putatively classified as F1, as their horizontal stratum lies near the lower margin of the FB. Their cell bodies form a cluster in the dorso-frontal cellular cortex above the antennal lobes. Like the F5 neurons, they have large arborizations in the dorsal unstructured neuropil. The line NP2320 expresses the driver in columnar neurons running perpendicular to the strata of F neurons, with their cell bodies scattered singly or in small groups between the calyces. As they seem to have no arborizations outside the FB, they are tentatively classified as pontine neurons³⁴.

First, we showed that pattern memory required the *rut* gene for each of the three parameters (data not shown). Next, we studied the Rut rescue flies (for example, *rut*²⁰⁸⁰/Y;c205/UAS-*rut*⁺). In the line c205, memory was restored only for 'elevation', not for 'size' or 'contour orientation' (Fig. 5a). Correspondingly, the memory impairment by expression of dominant-negative $G\alpha_s^*$ in this driver line should be specific for 'elevation', as is indeed the case (Fig. 4b). With the driver line NP6510, memory was not restored for either 'elevation' or for 'size', but memory was restored for 'contour orientation'. The third driver line, NP2320, labelling columnar neurons of the FB, did not restore the memory for any of the three

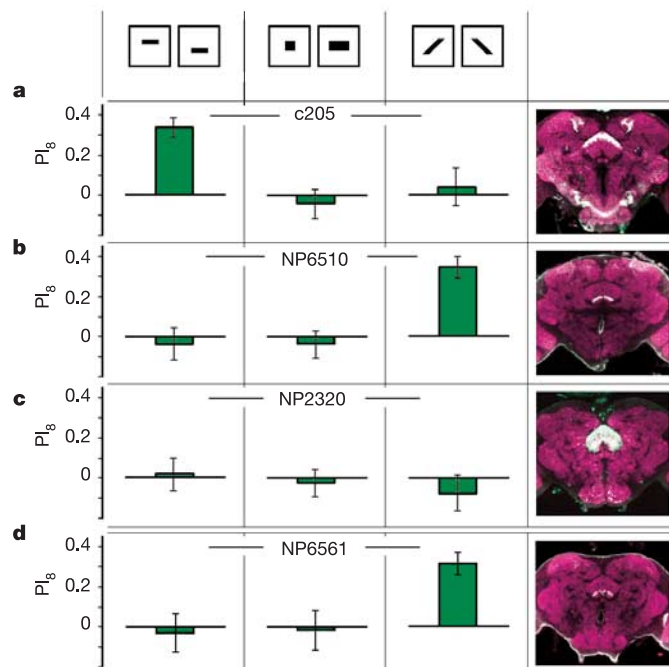


Figure 5 | Memory traces for pattern parameters are spatially separated.

a–c, Three GAL4 lines—c205 (**a**), NP6510 (**b**), NP2320 (**c**)—driving *rut*⁺ cDNA in a mutant *rut* background, were tested for 3-min memory using patterns specifically testing for the single parameters 'elevation', 'size' or 'contour orientation'. Expression patterns in the FB, using GFP as effector, are shown on the right (staining as in Fig. 3). Note that in NP2320 (**c**), only columnar elements (presumably 'pontine neurons')³⁴ are stained. **d**, Driver line NP6561 has a similar expression pattern as NP6510. In flies of the genotype *rut*²⁰⁸⁰/Y;NP6561/UAS-*rut*⁺, again only the memory for pattern parameter 'contour orientation' is rescued. Number of flies: *n* = 20 in all experiments. Error bars are s.e.m.

pattern parameters (Fig. 5c). Among the 27 GAL4 lines above (Supplementary Table S1) we found a second with a very similar expression pattern as NP6510 (NP6561). The P-element insertions in the two lines are only 124 nucleotides apart from each other. Like NP6510, NP6561 restores the memory for 'contour orientation' but not for 'size' or 'elevation' (Fig. 5d). These results strongly suggest that memory traces for distinct visual pattern parameters are located in different parts of the FB, and that, in addition to the memory trace in F5 neurons, a memory trace for the parameter 'contour orientation' is located in F1 neurons.

Developmental versus adult restoration

A pertinent question in rescue experiments is whether the rescue is due to the provision of an acute function in the adult or to the avoidance of a developmental defect. Therefore, the *tub*-GAL80^{ts} transposon was added to the system. We chose the driver lines c205 and NP6510. Groups of adult males (for example, *rut*²⁰⁸⁰/Y;+/+*tub*-GAL80^{ts};NP6510/UAS-*rut*⁺), raised at 19 °C, were kept as adults for 14 h at 19 °C or 30 °C. Afterwards, pattern memory for the corresponding pattern parameter was tested. In both cases, flies that had been kept at 30 °C showed normal memory, indicating that Rut cyclase induced just a few hours before the experiment had restored an immediate neuronal function rather than preventing a developmental defect (Fig. 4c). This conclusion was further supported by the finding that $G\alpha_s^*$ expression in the adult (using *tub*-GAL80^{ts}) was sufficient to disrupt memory (Fig. 4d).

Conclusions

Several conclusions can be drawn from the above results. Memory traces in *Drosophila* are associated with specific neuronal structures: odour memories with the MBs (reviewed in ref. 8), visual memories with the CX, and place memory (tentatively) with the median bundle²⁰. Memory traces are not stored in a common all-purpose memory centre. Even within the visual domain, memories for distinct pattern parameters are localized within distinct structures: a *rut*-dependent short-term memory trace for the pattern parameter 'elevation' to F5 neurons, and a corresponding memory trace for 'contour orientation' to F1 neurons. Moreover, if the constitutively activating $G\alpha_s^*$ protein indeed interferes with the regulation of Rut cyclase, it follows that the brain contains no other redundant *rut*-dependent memory traces for these pattern parameters. The Rut-mediated plasticity is necessary and sufficient, at least in F5 neurons. As in the earlier examples, the memory traces are confined to relatively small numbers of neurons. At least in flies, and probably in insects in general, memory traces appear to be part of the circuitry serving the respective behaviour.

Our study provides a first glimpse of the circuitry within a neural system for visual pattern recognition. Though the picture is far from complete, it invites (and may guide) speculation. The FB is a fibre matrix of layers, sectors and shells³⁴. The F1- and F5-neurons form two sharp parallel horizontal strata in this matrix. If the width of the FB represents the azimuth of visual space as recently proposed²⁴, the horizontal strata of the F neurons would be well suited to mediate translation invariance (see ref. 7 and the Supplementary Information). In any case, it is satisfying to find a translation invariant memory trace in the CX where visual information from both brain hemispheres converges^{35,22}. These first components of the circuitry may encourage modelling efforts for pattern recognition in small visual systems⁷ (see the Supplementary Information for a discussion of the pattern recognition mechanism).

METHODS

The apparatus for studying visual learning is shown in Fig. 1a (see also refs 36, 37). The tethered fly, suspended at a torque meter, is flying stationarily in the centre of an arena that is illuminated from behind and carries two upright and two inverted T-shaped patterns in alternating sequence on its wall. The arena is rotated such that its angular velocity is proportional to, but directed against, the

fly's yaw torque; that is, the fly's yaw torque determines the angular velocity of the arena instead of the fly's own body. This arrangement allows the tethered fly to stabilize and choose its flight orientation with respect to the arena by adjusting its yaw torque. A computer continuously registers yaw torque and angular position of the arena. For visual pattern learning, a beam of infrared light is directed at the fly as an instantaneous source of heat. The arena is virtually divided into four quadrants with the patterns at their respective centres (Fig. 1a; Supplementary Fig. S1). During training, the computer switches the heat beam on whenever the fly is heading towards a quadrant with, for example, an upright T, and switches it off when the fly is oriented towards one of the other two quadrants. Hence, half of all possible orientations in the arena are paired with heat, the other with ambient temperature. During tests, heat is permanently switched off. (In some experiments, a circular LED display is used instead of the mechanically rotating arena.)

Angular position is recorded every 50 ms and orientation preferences are calculated for nine consecutive 2-min periods (performance index (PI) 1–9). Pattern A is paired with ambient temperature during training and pattern B with heat. The two patterns alternate as A and B from fly to fly. If t_A is the time the fly spends heading towards the quadrants of pattern A, and t_B the time heading towards pattern B quadrants, the performance index is calculated as $PI = (t_A - t_B)/(t_A + t_B)$. Supplementary Fig. S1 gives a further explanation. Technical details regarding the flight simulator and data processing routines can be found in refs 36, 37. Error bars in the figures are s.e.m.; asterisks indicate levels of significance against zero (*** $P \leq 0.001$; ** $P \leq 0.01$; * $P \leq 0.05$; n.s., $P > 0.05$ (one-sample *t*-test; two-sided *P*-value)).

Details about the visual figures, fly maintenance, mutant lines, histological procedures and image processing are available in the Supplementary Information.

Received 27 May; accepted 25 October 2005.

- Heisenberg, M. & Wolf, R. On the fine structure of yaw torque in visual flight orientation of *Drosophila melanogaster*. *J. Comp. Physiol. [A]* **130**, 113–130 (1979).
- Xia, S., Liu, L., Feng, C. & Guo, A. Memory consolidation in *Drosophila* operant visual learning. *Learn. Mem.* **4**, 205–218 (1997).
- Dill, M. & Heisenberg, M. Visual pattern memory without shape recognition. *Phil. Trans. R. Soc. Lond. B* **349**, 143–152 (1995).
- Dill, M., Wolf, R. & Heisenberg, M. Visual pattern recognition in *Drosophila* involves retinotopic matching. *Nature* **365**, 751–753 (1993).
- Dill, M., Wolf, R. & Heisenberg, M. Behavioral analysis of *Drosophila* landmark learning in the flight simulator. *Learn. Mem.* **2**, 152–160 (1995).
- Ernst, R. & Heisenberg, M. The memory template in *Drosophila* pattern vision at the flight simulator. *Vision Res.* **39**, 3920–3933 (1999).
- Tang, S., Wolf, R., Xu, S. & Heisenberg, M. Visual pattern recognition in *Drosophila* is invariant for retinal position. *Science* **305**, 1020–1022 (2004).
- Gerber, B., Tanimoto, H. & Heisenberg, M. An engraving found? Evaluating the evidence from fruit flies. *Curr. Opin. Neurobiol.* **14**, 737–744 (2004).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- McGuire, S. E., Le, P. T., Osborn, A. J., Matsumoto, K. & Davis, R. L. Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765–1768 (2003).
- Dudai, Y., Corfas, G. & Hazvi, S. What is the possible contribution of Ca^{2+} -stimulated adenylate cyclase to acquisition, consolidation and retention of an associative olfactory memory in *Drosophila*. *J. Comp. Physiol. [A]* **162**, 101–109 (1988).
- Levin, L. R. et al. The *Drosophila* learning and memory gene *rutabaga* encodes a Ca^{2+} /Calmodulin-responsive adenylyl cyclase. *Cell* **68**, 479–489 (1992).
- Abrams, T. W., Yovell, Y., Onyike, C. U., Cohen, J. E. & Jarrard, H. E. Analysis of sequence-dependent interactions between transient calcium and transmitter stimuli in activating adenylyl cyclase in *Aplysia*: possible contribution to CS-US sequence requirement during conditioning. *Learn. Mem.* **4**, 496–509 (1998).
- Renger, J. J., Ueda, A., Atwood, H. L., Govind, C. K. & Wu, C. F. Role of cAMP cascade in synaptic stability and plasticity: ultrastructural and physiological analyses of individual synaptic boutons in *Drosophila* memory mutants. *J. Neurosci.* **20**, 3980–3992 (2000).
- Zars, T., Fischer, M., Schulz, R. & Heisenberg, M. Localization of a short-term memory in *Drosophila*. *Science* **288**, 672–675 (2000).
- McGuire, S. E., Le, P. T. & Davis, R. L. The role of *Drosophila* mushroom body signalling in olfactory memory. *Science* **293**, 1330–1333 (2001).
- Connolly, J. B. et al. Associative learning disrupted by impaired G_s signalling in *Drosophila* mushroom bodies. *Science* **274**, 2104–2107 (1996).
- Dubnau, J., Grady, L., Kitamoto, T. & Tully, T. Disruption of neurotransmission in *Drosophila* mushroom body blocks retrieval but not acquisition of memory. *Nature* **411**, 476–480 (2001).
- Schwaerzel, M., Heisenberg, M. & Zars, T. Extinction antagonizes olfactory memory at the subcellular level. *Neuron* **35**, 951–960 (2002).

20. Zars, T., Wolf, R., Davis, R. & Heisenberg, M. Tissue-specific expression of a type I adenylyl cyclase rescues the *rutabaga* mutant memory defect: in search of the engram. *Learn. Mem.* **7**, 18–31 (2000).
21. Heisenberg, M. *Central Brain Function in Insects: Genetic Studies on the Mushroom Bodies and Central Complex in Drosophila* (eds Schildberger, K. & Elsner, N.) (Gustav Fischer, Stuttgart/Jena/New York, 1994).
22. Strauss, R. The central complex and the genetic dissection of locomotor behaviour. *Curr. Opin. Neurobiol.* **12**, 633–638 (2002).
23. Homberg, U. In search of the sky compass in the insect brain. *Naturwissenschaften* **91**, 199–208 (2004).
24. Strauss, R. *Die übergeordnete Steuerung des Laufverhaltens durch das Insektengehirn, studiert mit Methoden der Drosophila-Neurogenetik*. Habilitation thesis, Bayerische Julius-Maximilians-Universität, Würzburg (2002).
25. Weidtmann, N. *Visuelle Flugsteuerung und Verhaltensplastizität bei Zentralkomplexmutanten von Drosophila melanogaster*. Diploma thesis, Bayerische Julius-Maximilians-Universität, Würzburg (1993).
26. Strauss, R. & Heisenberg, M. A higher control center of locomotor behaviour in the *Drosophila* brain. *J. Neurosci.* **13**, 1852–1861 (1993).
27. Keller, A., Sweeney, S. T., Zars, T., O'Kane, C. J. & Heisenberg, M. Targeted expression of tetanus neurotoxin interferes with behavioural responses to sensory input in *Drosophila*. *J. Neurobiol.* **50**, 221–233 (2002).
28. Broadie, K. *et al.* Syntaxin and synaptobrevin function downstream of vesicle docking in *Drosophila*. *Neuron* **15**, 663–673 (1995).
29. Eyding, D. *Lernen und Kurzzeitgedächtnis beim operanten Konditionieren auf visuelle Muster bei strukturellen und biochemischen Mutanten von Drosophila melanogaster*. Diploma thesis, Bayerische Julius-Maximilians-Universität, Würzburg (1993).
30. Brembs, B. & Heisenberg, M. The operant and the classical in conditioned orientation of *Drosophila melanogaster* at the flight simulator. *Learn. Mem.* **7**, 104–115 (2000).
31. Prokop, A. & Technau, G. M. Normal function of the mushroom body defect gene of *Drosophila* is required for the regulation of the number and proliferation of neuroblasts. *Dev. Biol.* **161**, 321–337 (1994).
32. Wolf, R. *et al.* *Drosophila* mushroom bodies are dispensable for visual, tactile, and motor learning. *Learn. Mem.* **5**, 166–178 (1998).
33. Fischbach, K. F. & Dittrich, A. P. M. The optic lobe of *Drosophila melanogaster*. A Golgi analysis of wild-type structure. *Cell Tissue Res.* **258**, 441–475 (1989).
34. Hanesch, U., Fischbach, K. F. & Heisenberg, M. Neuronal architecture of the central complex in *Drosophila melanogaster*. *Cell Tissue Res.* **257**, 343–366 (1989).
35. Vitzthum, H., Müller, M. & Homberg, U. Neurons of the central complex of the locust *Schistocerca gregaria* are sensitive to polarized light. *J. Neurosci.* **22**, 1114–1125 (2002).
36. Wolf, R. & Heisenberg, M. Basic organization of operant behaviour as revealed in *Drosophila* flight orientation. *J. Comp. Physiol. [A]* **169**, 699–705 (1991).
37. Heisenberg, M. & Wolf, R. Reafferent control of optomotor yaw torque in *Drosophila melanogaster*. *J. Comp. Physiol. [A]* **163**, 373–388 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Gerber and H. Tanimoto for valuable comments on the manuscript, C. Grübel, Haiyun Gong and Huoqing Jiang for excellent technical assistance, and A. Jenett for visualizing gene expression patterns. Supported by Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie (M.H.), National Natural Sciences Foundation of China (L.L.), '973-program' (L.L.), and by the Knowledge Innovation Project of the Chinese Academy of Sciences (L.L.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.L. (liuli@sun5.ibp.ac.cn) or M.H. (heisenberg@biozentrum.uni-wuerzburg.de).

Replication fork reactivation downstream of a blocked nascent leading strand

Ryan C. Heller¹ & Kenneth J. Marians^{1,2}

Unrepaired lesions in the DNA template pose a threat to accurate replication. Several pathways exist in *Escherichia coli* to reactivate a blocked replication fork. The process of recombination-dependent restart of broken forks is well understood, but the consequence of replication through strand-specific lesions is less well known. Here we show that replication can be restarted and leading-strand synthesis re-initiated downstream of an unrepaired block to leading-strand progression, even when the 3'-OH of the nascent leading strand is unavailable. We demonstrate that the loading by a replication restart system of a single hexamer of the replication fork helicase, DnaB, on the lagging-strand template is sufficient to coordinate priming by the DnaG primase of both the leading and lagging strands. These observations provide a mechanism for damage bypass during fork reactivation, demonstrate how daughter-strand gaps are generated opposite leading-strand lesions during the replication of ultraviolet-light-irradiated DNA, and help to explain the remarkable speed at which even a heavily damaged DNA template is replicated.

DNA template damage arises from a variety of sources and can have many adverse effects on the genomic integrity of the cell. Left unrepaired, DNA damage can lead to genetic mutation, chromosomal rearrangements¹ and cancer. In early studies on the effects of a damaged template on DNA replication in *E. coli*, cells were irradiated with ultraviolet light, which causes photoreversible pyrimidine dimers to form on the DNA template². Strains defective in nucleotide excision repair were used, so that the damaged bases would not be removed before encounter with the replication fork. Unexpectedly, even with a dose of ultraviolet light high enough to induce 270 pyrimidine dimers in the genome, DNA synthesis was not halted³. Instead, the replication rate was reduced, with a relatively brief delay of approximately 10 s per dimer. Analysis of the DNA after irradiation revealed that daughter-strand gaps were present in both nascent strands and were about 1,000 nucleotides in length⁴. Because the daughter-strand gaps were located opposite the pyrimidine dimers⁵, this suggested that the presence of the dimer in the template was causing DNA synthesis to stop, but that it was somehow able to restart at some point beyond the dimer. That the pyrimidine dimers remained in the DNA after daughter-strand gap repair⁶ and that mutations continued to arise more than one generation after treatment with ultraviolet light⁷ suggested that dimer excision was not a necessary prerequisite for the continuation of DNA replication.

Owing to the asymmetric nature of DNA replication, a collision between the replisome and a DNA lesion can have different consequences depending on which strand is damaged. During replication *in vitro* of plasmid molecules containing a blocking lesion located in the lagging-strand template, replication fork progression is not compromised, and so there is no need for restart under these circumstances^{8,9}. Synthesis of the Okazaki fragment is blocked, but the two polymerases remain physically associated, allowing leading-strand synthesis to continue. Lagging-strand synthesis simply resumes after the next cycle of Okazaki fragment priming, leaving a small daughter-strand gap in the nascent DNA.

However, if the block is located in the leading-strand template, the polymerases can become uncoupled, and continued template

unwinding allows nascent lagging-strand synthesis to proceed past the blocked nascent leading strand. The resulting stalled fork contains a single-stranded region on the leading strand between the lesion and the fork duplex that was estimated to be about 1 kb *in vitro*⁸ and at least 1 kb in an *in vivo* study¹⁰. Several models have been proposed to explain how a replication fork that has stalled from a leading-strand-specific block could be restarted in order to complete replication of the chromosome. If the nascent strands anneal with each other, the stalled fork could undergo regression to form a four-way, Holliday junction^{11,12}. From this point, one possibility is that the nascent leading strand is extended, using the nascent lagging strand as a template, a process called template switching. The regressed fork would then be reversed, and PriA-dependent replisome assembly¹³ would restart replication from the fork structure. Another possibility is that the blocking lesion is removed while the fork is regressed, and the nascent DNA is degraded to reset the Holliday junction to a fork structure for restart¹⁴. Alternatively, the regressed fork could undergo RuvABC-directed Holliday-junction branch migration and cleavage. The recombination proteins would act on the broken chromosome end, and then undergo a strand-invasion of the chromosome to form a D-loop. Once the original block is removed, PriA could then direct replisome assembly for the completion of chromosomal replication. Although the processing of the stalled fork through a Holliday junction intermediate generates structures amenable to restart for each of the proposed models, some models conflict with the *in vivo* evidence that ultraviolet-light-induced pyrimidine dimers can remain in the template after passage of the replication fork⁶. One fundamental problem with all of the proposed models is that none account for the formation of gaps in the leading strand after the encounter of the replication fork with the template lesion. In each case, replication fork reactivation would require extension of the nascent leading-strand DNA as a primer for continued leading-strand synthesis.

Repriming of the nascent leading strand

On the basis of previous work, the structure of a fork stalled by a leading-strand block (Fig. 1a) is an excellent substrate for replication

¹Molecular Biology Program, Weill Graduate School of Medical Sciences of Cornell University and ²Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

restart by the PriC-dependent DnaB loading/restart system because of the large gap on the leading-strand template adjacent to the fork junction¹⁵. Using a 6.9-kb linear, forked template DNA carrying a gap in the nascent leading strand, we have previously shown that a reconstituted PriC-dependent restart reaction is partially dependent on DnaG, even for leading-strand synthesis, suggesting that the leading strand was being re-initiated and that the oligonucleotide representing the nascent leading strand was not elongated in some cases¹⁵. This is in contrast to the PriA system and a previous study in which leading-strand synthesis was completely independent of the primase¹⁶. To directly address the question of whether DNA synthesis could be primed on the leading strand during restart, we performed a

PriC-dependent restart reaction on a forked, linear template DNA carrying no oligonucleotides (Fig. 1b), representing a stalled fork with a large leading-strand gap. On this single-stranded DNA-binding protein (SSB)-coated fork, PriC loads DnaB to the lagging-strand template, which is sufficient to allow replisome assembly and unwinding in the 5' → 3' direction. Even though there was no nascent leading-strand oligonucleotide to extend, we observed both leading- and lagging-strand products (lane 6), which must have been initiated by priming *de novo* on their respective template strands. Replication is completely dependent on all PriC-system components, including DnaG (lanes 1–5). These results support the idea that loading DnaB to an unprimed fork is not an abortive or deleterious

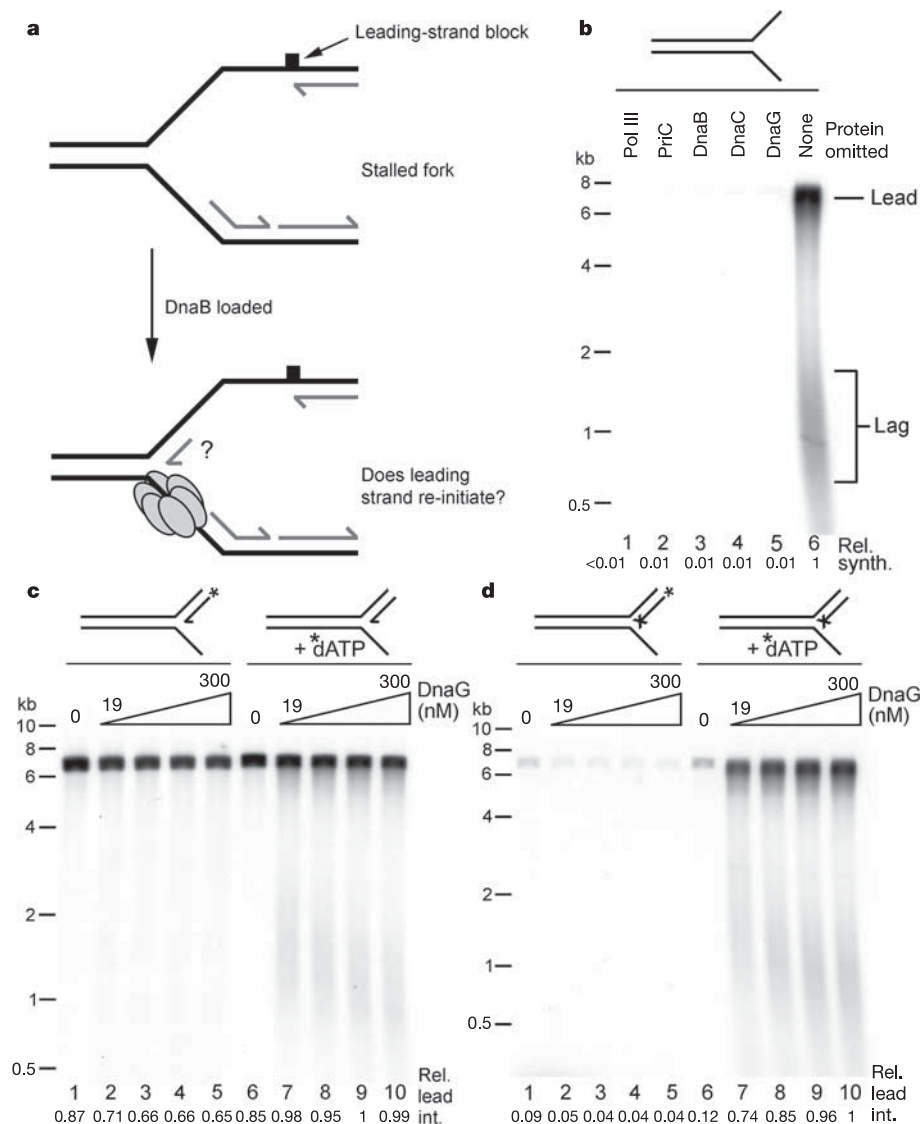


Figure 1 | The leading strand can be reprimed during both PriA- and PriC-dependent restart. **a**, A stalled fork in which lagging-strand synthesis has progressed past the nascent leading strand is an excellent substrate for PriC-dependent restart. Arrowheads indicate the direction of 5' → 3' DNA synthesis. DnaB is represented by the grey ovals. **b**, A complete PriC restart reaction using the linear fork template with no oligonucleotides as substrate. Protein components are omitted as indicated. A relative DNA synthesis value of 1 represents 9.8 pmol of [³²P]dAMP incorporated into acid-insoluble product. **c**, **d**, Complete PriA restart reactions using the linear fork template carrying a 5'-radiolabelled leading strand as the only source of label (indicated by an asterisk on the oligonucleotide), or an unlabelled oligonucleotide and [³²P]dATP. The quantity of [³²P]dATP was adjusted such that the specific activity of a fully replicated nascent leading

strand matched that of the labelled oligonucleotide. DnaG, where present, increases 2.5-fold from left to right. The leading-strand oligonucleotide is either unmodified (**c**) or contains a 3'-terminal 2',3'-dideoxy cytosine (**d**). A maximum of 13.1 pmol and 9.6 pmol [³²P]dAMP was incorporated into acid-insoluble product in **c**, and **d**, respectively. The ratio of leading- to lagging-strand products was measured to be 1.15 (**c**, lane 10) and 1.09 (**d**, lane 10). The slight elongation of the leading-strand oligonucleotide that is observed in lane 1 of panel **d**, probably results from 3' → 5' exonucleolytic removal of the 3'-dideoxy nucleotide. The reaction uses 16 different preparations of purified proteins. It is very difficult, in the aggregate, to ensure complete removal of all trace contamination of nucleases. Rel. lead. int., relative intensity of leading-strand signal.

event, but may be an efficient way to bypass damage and continue replication of the chromosome with minimal delay.

We next determined whether the PriA system was similarly capable of priming leading-strand synthesis. This experiment requires a fork carrying an oligonucleotide representing the nascent leading-strand DNA for efficient recognition by PriA¹⁵. If the oligonucleotide was 5'-end-labelled and the fork was reacted in the absence of DnaG, full-length leading-strand product DNA was observed (Fig. 1c, lane 1). In the presence of increasing concentrations of DnaG, the leading-strand signal shows a slight decrease in intensity, but increases in intensity when all nascent DNA is labelled (compare lanes 1–5 with lanes 6–10), suggesting that some leading-strand product may be re-initiated even when the nascent strand is unblocked. However, because most of the signal is present when the oligonucleotide is the only source of label, we conclude that the nascent leading strand is frequently extended during PriA-dependent restart, even in the presence of high concentrations of DnaG.

We replaced the leading-strand oligonucleotide with one containing a chain-terminating 2',3'-dideoxy cytosine to test whether the leading strand is re-initiated downstream of a block during PriA-dependent restart (Fig. 1d). As expected, replicating from the fork containing the 5'-labelled dideoxy-terminating oligonucleotide

(lanes 1–5) produced very little signal because elongation by DNA polymerase III holoenzyme (Pol III HE) lacking the proofreading subunit could not occur. In actuality, the substrate was replicated extensively in the presence of DnaG, but the products were only revealed when all nascent DNA was labelled by the inclusion of labelled dATP (lanes 6–10), indicating that as with the PriC system, DnaG can prime both the leading and lagging strands. Leading-strand priming during PriA-dependent restart may occur on those forks that are stalled by a leading-strand lesion but that fail to undergo polymerase uncoupling, and therefore do not contain a gap produced by the ensuing template unwinding. Alternatively, a RecA filament may present an impediment to the extension of the invading strand in a D-loop, but not to DnaB loading, leading-strand priming or replication restart. In this case, repriming of the leading strand could provide an efficient and rapid means to continue replication even before RecA disassembly from the recombination joint.

In support of this premise, it has previously been shown that on a D-loop-containing joint molecule generated by the combined actions of RecA and RecBCD, PriA-dependent restart occurred but was nearly completely DnaG-dependent¹⁷, suggesting that a majority of the nascent leading strand was re-initiated.

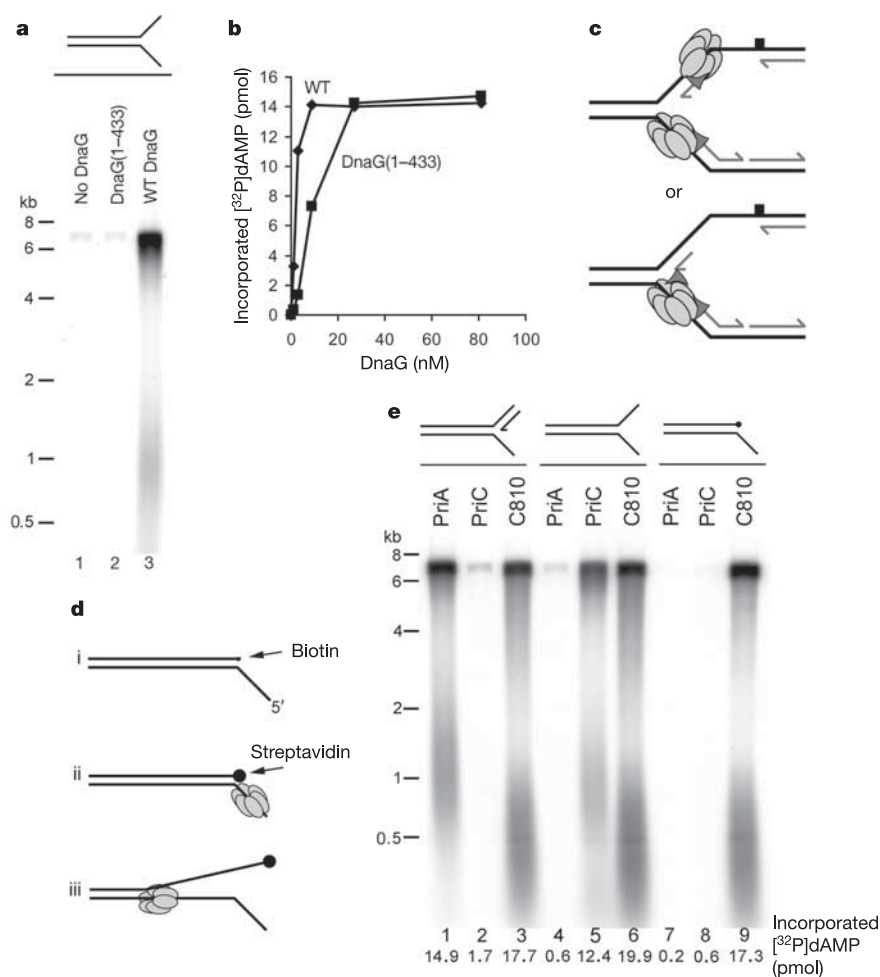


Figure 2 | A modified linear template in which the fork 3'-arm is replaced with a biotin group. **a**, Complete PriC-dependent replication restart reaction using a fork template carrying no oligonucleotide with either no DnaG, a DnaG truncation mutant containing the N-terminal 433 amino acids, or wild-type DnaG. **b**, The DnaG mutant retains primase activity in an M13Gori1 replication assay. In this assay, primase synthesizes a specific 28-nt-long primer at a hairpin region on the viral DNA that is then elongated by the Pol III HE in the presence of SSB. There is no DnaB present. **c**, Either

two DnaB hexamers are loaded to a fork during restart, or DnaG (grey triangle) interacts with the DnaB on the lagging-strand template for priming both strands. **d**, On the modified template, DnaB binds to the single-stranded DNA representing the fork lagging-strand template. Unwinding can occur only when streptavidin is bound. **e**, Complete PriA, PriC, and DnaC810 bypass replication restart reactions using the fork carrying a leading-strand oligonucleotide, fork template alone, or the modified biotin template. Each reaction contains 12 nM streptavidin. WT, wild-type DnaG.

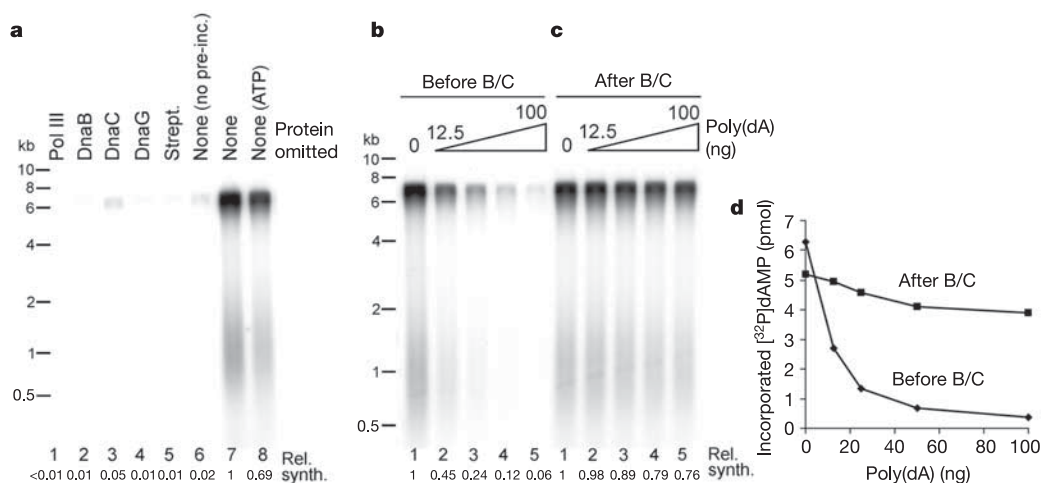


Figure 3 | A single DnaB hexamer on the lagging-strand template coordinates priming of both strands. **a**, Complete two-step replication of the modified biotin template. Components are omitted or all components are added simultaneously, as indicated. Where noted, the reaction mixture contained 5 μ M ATP in the preincubation step. **b**, **c**, Two-step reaction of linear biotin template. Poly(dA), where present, increases twofold from left

to right. Competitor was either added before (**b**) or after (**c**) the preincubation step, then incubated at 37 °C for an additional 5 min before the remainder of the replication components were added. Before and after B/C, before and after addition of DnaB and DnaC. Rel. synth., relative DNA synthesis. **d**, A plot showing the amount of [³²P]dAMP incorporated into acid-insoluble product for the reactions in **b** and **c**.

One DnaB hexamer coordinates priming

The DnaG-directed leading-strand priming that we observed was probably occurring through interaction with a DnaB hexamer, because it has previously been shown that DnaG alone is nearly inactive with respect to synthesizing primers on SSB-coated templates¹⁸, but is targeted to the replication fork by a protein–protein interaction with DnaB¹⁹. To confirm that the DnaG–DnaB interaction was essential for leading-strand priming, we tested a truncated DnaG protein (containing only the amino-terminal 433 amino acids) that is incapable of interaction with DnaB¹⁹. Indeed, the mutant DnaG was inactive with respect to sustaining both leading- and lagging-strand synthesis (Fig. 2a), even though the protein retained the ability to synthesize a primer (Fig. 2b). Is the productive interaction for leading-strand synthesis an Okazaki fragment-type priming event with a second DnaB on the leading-strand template, or does the single DnaB on the lagging-strand template interact with a second DnaG and undergo a new mode of priming on the leading strand (Fig. 2c)?

To address the latter mechanism, we modified the linear fork template by replacing the 3' arm of the fork with a biotin group (Fig. 2d). The advantage of this template is that it contains only one region of single-stranded DNA onto which DnaB can be initially loaded, and thus priming of the leading strand can only occur after unwinding is initiated. The unwinding of a duplex by DnaB typically requires that an unpaired 3' arm be excluded from the centre of the helicase. To overcome that requirement, replication is initiated in the presence of streptavidin, which has a 54 Å diameter and is too large to fit in the centre of the helicase. When bound to the biotin group at the single strand–double strand junction, streptavidin will be displaced from the centre of the helicase, allowing unwinding of the duplex²⁰. In the absence of streptavidin, the entire DNA duplex would enter the centre channel of the helicase²¹, which would be nonproductive for unwinding and replication.

This linear biotin template no longer has a fork structure, so the PriA and PriC loading systems are not active, even in the presence of streptavidin (Fig. 2e). The template was, however, capable of sustaining replication in the presence of the gain-of-function mutant DnaC810, which can load DnaB to any SSB-coated DNA²².

Because it is possible that DnaC810 loads a second DnaB to the leading-strand template during unwinding, and that the PriA and PriC loading systems are not active on this template, we used a two-step reaction to determine whether a single DnaB hexamer on the

lagging-strand template was sufficient to allow leading-strand priming. First, DnaB was pre-loaded onto the DNA with wild-type DnaC in the absence of SSB. Upon further incubation with nucleotides, streptavidin, Pol III HE, DnaG and SSB (which prevents another DnaB from binding to the leading-strand template), an approximately equal proportion of leading- and lagging-strand products were observed, indicating efficient leading-strand priming (Fig. 3a, lane 7). Each of the proteins was required (lanes 1–5), as was the preincubation step (lane 6). Because DnaC stably associates with the DnaB hexamer²³ and is fully capable of loading DnaB to single-stranded DNA in the presence of ADP²⁴, we used this nucleoside diphosphate in the pre-incubation step to prevent helicase movement and maximize stability and product accumulation, although a low concentration of ATP could be substituted with little negative effect (lane 8).

To rule out the possibility that free DnaB gained access to the fork during unwinding and was responsible for the observed leading-strand priming, we challenged the reaction with the polydeoxyribonucleotide poly(dA), which binds DnaB with high affinity²⁵, in order to sequester the unbound DnaB. If the competitor was added to the reaction before DnaB was loaded, replication activity was nearly completely inhibited (Fig. 3b, d). However, the reaction was relatively insensitive to poly(dA) if challenged after a DnaB hexamer was bound to the single-stranded DNA (Fig. 3c, d), demonstrating that the same hexameric helicase that unwinds the replication fork duplex was also responsible for interacting with DnaG to prime leading-strand synthesis as well as the lagging strand for Okazaki fragment synthesis. That multiple primase monomers can bind a single DnaB hexamer²⁶ supports the idea of this multifunctional role for DnaB at the replication fork.

Discussion

Our data indicate that although a blocking lesion located in the leading strand may stall the replication fork temporarily, the lesion need not be repaired, and neither nascent strand regression nor fork breakage is required in order to resume DNA synthesis (Fig. 4). PriC-dependent loading of DnaB is sufficient to reassemble the replisome and coordinate the priming of both the leading and lagging strands. On the basis of this model, it can be predicted that repriming the leading strand downstream of a block will leave a gap opposite the lesion, one that cannot be filled in simply, except by specialized translesion polymerases. This prediction is substantiated

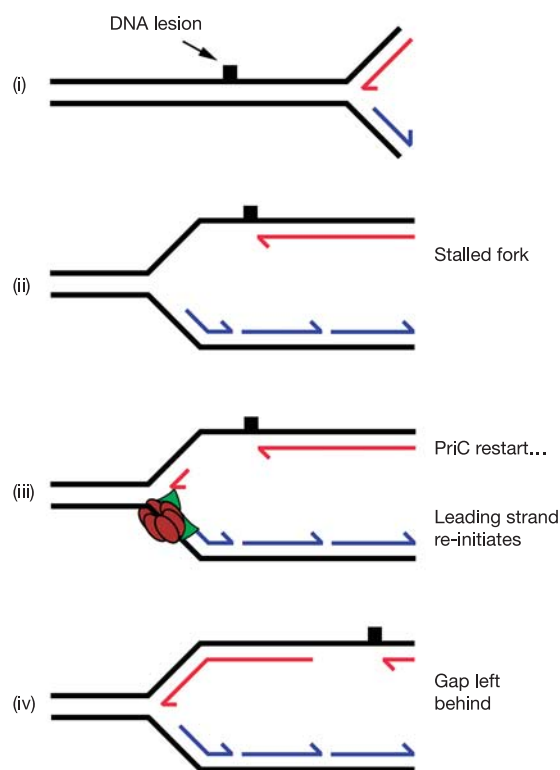


Figure 4 | PriC-dependent restart of a stalled fork generates daughter-strand gaps. When the replisome encounters a lesion in the leading-strand template, unwinding continues for some time, so the stalled fork contains an exposed region of single-stranded DNA. During the process of replication restart by PriC, the leading strand is re-initiated *de novo*, leaving a daughter-strand gap opposite the lesion. Brown ovals, DnaB; green triangle, DnaG.

by studies *in vivo*, which have shown that upon ultraviolet-light-irradiation of strains defective in nucleotide excision repair, DNA synthesis continued, and daughter-strand gaps of about 1,000 nucleotides in length⁴ were produced in both daughter strands. These gaps are located opposite the dimers⁵, and are repaired primarily by a recombinational exchange process involving RecA, although a minority are filled by the mutagenic polV, perhaps in situations where daughter strand gaps overlap²⁷. In the absence of PriA, when PriC-dependent restart is the only available pathway, gap-repair by RecFOR becomes essential²⁸, further associating gap formation with replication restart. Disrupting both *priA* and *priC* is lethal²⁹, indicating that essentially all forks initiating from *oriC* undergo stalling and replication restart at least once, and so daughter-strand gap formation may also be prevalent in the absence of exogenous damaging agents, depending on the relative contributions of strand-specific blocks during normal cell growth. Daughter-strand gap formation after DNA damage induction has been detected in a variety of organisms, including yeast³⁰, mice³¹ and human cells³², suggesting that polymerase uncoupling, and therefore by association, replication restart, are universal mechanisms for the bypass of certain DNA damage.

During Okazaki's classic studies of DNA replication mechanisms *in vivo*, it was found that all label during short pulse-labelling of nascent DNA in *E. coli* was present as short chains during alkaline sucrose sedimentation³³. A significant number of reports along these lines led to models in which both nascent strands were replicated in a discontinuous fashion³⁴. In contrast to the work *in vivo*, biochemical reconstitutions of DNA replication have clearly demonstrated that the leading strand is synthesized in a mechanistically continuous fashion, a disparity that has never been satisfactorily resolved. The finding in this study that the nascent leading strand can be reprimed

outside *oriC* during replication restart might provide a mechanism for the effective discontinuity observed *in vivo*.

METHODS

Proteins and DNA. Replication restart proteins were prepared as previously described¹⁵. Recombinant streptavidin was purchased from Roche. The linear forked template DNA and the template containing leading-strand oligonucleotide were constructed as previously described¹⁵. The leading-strand oligonucleotide with the 3'-terminal 2',3'-dideoxy cytosine has the same sequence as the unmodified oligonucleotide and was synthesized by Sigma-Genosys. The biotinylated template was constructed as described¹⁵, except that an oligonucleotide with the sequence 5'-GGCCGTGAGACCGCAATACGGATAAGGGCT-biotin-3' was substituted for the 68-nt oligonucleotide that forms the fork 3' arm.

Replication restart reactions. Unless otherwise indicated, reaction mixtures (15 μ l) contained the standard components 50 mM HEPES-KOH (pH 8.0), 12 mM MgOAc₂, 10 mM dithiothreitol, 100 μ g ml⁻¹ bovine serum albumin, 1 mM ATP, 200 μ M CTP, UTP and GTP, 40 μ M [α -³²P]dATP (4,000–5000 cpm pmol⁻¹), 40 μ M dCTP, dTTP and dGTP, 1 nM linear template DNA, 250 nM SSB, 20 nM Pol III HE, 60 nM DnaB, 400 nM DnaC and 300 nM DnaG. For PriC-dependent restart, reactions additionally contained 320 nM PriC. For PriA-dependent restart, reactions also contained 20 nM PriA, 40 nM PriB and 480 nM DnaT, and the ATP concentration was 0.2 mM. For bypass reactions, DnaC was replaced with the DnaC810 mutant. For two-step replication of the linear biotinylated template, reaction mixtures contained template DNA, DnaB, DnaC, HEPES-KOH (pH 8.0), MgOAc₂, bovine serum albumin and 500 μ M ADP, and were preincubated at 37 °C for 5 min, followed by the addition of the remainder of the standard components plus 12 nM streptavidin. All reactions were incubated at 37 °C for 20 min and DNA synthesis was terminated by the addition of EDTA to a concentration of 33 mM. Analysis of products and quantification of total DNA synthesis was as described¹⁵. Leading-strand signal was quantified by Fuji ImageGauge software after phosphorimager scan.

M13Gor1 replication. Reaction mixtures (25 μ l) contained the following components: 50 mM HEPES-KOH (pH 8.0), 10 mM MgOAc₂, 10 mM dithiothreitol, 200 μ g ml⁻¹ bovine serum albumin, 1 mM ATP, 200 μ M CTP, UTP and GTP, 40 μ M [α -³²P]dATP (1,300 cpm pmol⁻¹), 40 μ M dCTP, dTTP and dGTP, 2 nM single-stranded circular M13Gor1 DNA, 1.62 μ M SSB, 20 nM Pol III HE, and the indicated concentrations of DnaG or DnaG(1–433). Incubation was at 30 °C for 5 min, followed by the addition of 0.1 ml of NaPP_i; and determination of acid-insoluble radioactivity.

Received 1 August; accepted 17 October 2005.

- Myung, K., Chen, C. & Kolodner, R. D. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* **411**, 1073–1076 (2001).
- Witkin, E. M. Ultraviolet-induced mutation and DNA repair. *Annu. Rev. Microbiol.* **23**, 487–514 (1969).
- Rupp, W. D. & Howard-Flanders, P. Discontinuities in the DNA synthesized in an excision-defective strain of *Escherichia coli* following ultraviolet irradiation. *J. Mol. Biol.* **31**, 291–304 (1968).
- Iyer, V. N. & Rupp, W. D. Usefulness of benzoylated naphthoylated DEAE-cellulose to distinguish and fractionate double-stranded DNA bearing different extents of single-stranded regions. *Biochim. Biophys. Acta* **228**, 117–126 (1971).
- Bridges, B. A. & Sedgwick, S. G. Effect of photoreactivation on the filling of gaps in deoxyribonucleic acid synthesized after exposure of *Escherichia coli* to ultraviolet light. *J. Bacteriol.* **117**, 1077–1081 (1974).
- Ganesan, A. K. Persistence of pyrimidine dimers during post-replication repair in ultraviolet light-irradiated *Escherichia coli* K12. *J. Mol. Biol.* **87**, 103–119 (1974).
- Bridges, B. A. & Munson, R. J. Mutagenesis in *Escherichia coli*: evidence for the mechanism of base change mutation by ultraviolet radiation in a strain deficient in excision-repair. *Proc. R. Soc. Lond. B* **171**, 213–226 (1968).
- Higuchi, K. *et al.* Fate of DNA replication fork encountering a single DNA lesion during *oriC* plasmid DNA replication *in vitro*. *Genes Cells* **8**, 437–449 (2003).
- McInerney, P. & O'Donnell, M. Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. *J. Biol. Chem.* **279**, 21543–21551 (2004).
- Pages, V. & Fuchs, R. P. Uncoupling of leading- and lagging-strand DNA replication during lesion bypass *in vivo*. *Science* **300**, 1300–1303 (2003).
- Higgins, N. P., Kato, K. & Strauss, B. A model for replication repair in mammalian cells. *J. Mol. Biol.* **101**, 417–425 (1976).
- Seigneur, M., Bidnenko, V., Ehrlich, S. D. & Michel, B. RuvAB acts at arrested replication forks. *Cell* **95**, 419–430 (1998).
- Sandler, S. J. & Mariani, K. J. Role of PriA in replication fork reactivation in *Escherichia coli*. *J. Bacteriol.* **182**, 9–13 (2000).
- Michel, B., Grompone, G., Flores, M. J. & Bidnenko, V. Multiple pathways process stalled replication forks. *Proc. Natl Acad. Sci. USA* **101**, 12783–12788 (2004).

15. Heller, R. C. & Marians, K. J. The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. *Mol. Cell* **17**, 733–743 (2005).
16. Wu, C. A., Zechner, E. L. & Marians, K. J. Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. *J. Biol. Chem.* **267**, 4030–4044 (1992).
17. Xu, L. & Marians, K. J. PriA mediates DNA replication pathway choice at recombination intermediates. *Mol. Cell* **11**, 817–826 (2003).
18. Swart, J. R. & Griep, M. A. Primase from *Escherichia coli* primes single-stranded templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins. Template sequence requirements based on the bacteriophage G4 complementary strand origin and Okazaki fragment initiation sites. *J. Biol. Chem.* **268**, 12970–12976 (1993).
19. Tougu, K., Peng, H. & Marians, K. J. Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork. *J. Biol. Chem.* **269**, 4675–4682 (1994).
20. Hacker, K. J. & Johnson, K. A. A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding. *Biochemistry* **36**, 14080–14087 (1997).
21. Kaplan, D. L. & O'Donnell, M. DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands. *Mol. Cell* **10**, 647–657 (2002).
22. Xu, L. & Marians, K. J. Purification and characterization of DnaC810, a primosomal protein capable of bypassing PriA function. *J. Biol. Chem.* **275**, 8196–8205 (2000).
23. Galletto, R., Jezewska, M. J. & Bujalowski, W. Interactions of the *Escherichia coli* DnaB helicase hexamer with the replication factor the DnaC protein. Effect of nucleotide cofactors and the ssDNA on protein-protein interactions and the topology of the complex. *J. Mol. Biol.* **329**, 441–465 (2003).
24. Davey, M. J., Fang, L., McInerney, P., Georgescu, R. E. & O'Donnell, M. The DnaC helicase loader is a dual ATP/ADP switch protein. *EMBO J.* **21**, 3148–3159 (2002).
25. Reha-Krantz, L. J. & Hurwitz, J. The dnaB gene product of *Escherichia coli*. II. Single stranded DNA-dependent ribonucleoside triphosphatase activity. *J. Biol. Chem.* **253**, 4051–4057 (1978).
26. Mitkova, A. V., Khopde, S. M. & Biswas, S. B. Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of *Escherichia coli* in RNA primer synthesis. *J. Biol. Chem.* **278**, 52253–52261 (2003).
27. Smith, K. C. Recombinational DNA repair: the ignored repair systems. *Bioessays* **26**, 1322–1326 (2004).
28. Grompone, G., Sanchez, N., Dusko Ehrlich, S. & Michel, B. Requirement for RecFOR-mediated recombination in priA mutant. *Mol. Microbiol.* **52**, 551–562 (2004).
29. Sandler, S. J. Multiple genetic pathways for restarting DNA replication forks in *Escherichia coli* K-12. *Genetics* **155**, 487–497 (2000).
30. Prakash, L. Characterization of postreplication repair in *Saccharomyces cerevisiae* and effects of rad6, rad18, rev3 and rad52 mutations. *Mol. Gen. Genet.* **184**, 471–478 (1981).
31. Lehmann, A. R. Postreplication repair of DNA in ultraviolet-irradiated mammalian cells. *J. Mol. Biol.* **66**, 319–337 (1972).
32. Meneghini, R. Gaps in DNA synthesized by ultraviolet light-irradiated WI38 human cells. *Biochim. Biophys. Acta* **425**, 419–427 (1976).
33. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl Acad. Sci. USA* **59**, 598–605 (1968).
34. Ogawa, T. & Okazaki, T. Discontinuous DNA replication. *Annu. Rev. Biochem.* **49**, 421–457 (1980).

Acknowledgements These studies were supported by a grant from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.J.M. (kmarians@sloankettering.edu).

The trans-neptunian object UB₃₁₃ is larger than Pluto

F. Bertoldi^{1,2}, W. Altenhoff², A. Weiss², K.M. Menten² & C. Thum³

The most distant known object in the Solar System, 2003 UB₃₁₃ (97 AU from the Sun), was recently discovered near its aphelion¹. Its high eccentricity and inclination to the ecliptic plane, along with its perihelion near the orbit of Neptune, identify it as a member of the 'scattered disk'. This disk of bodies probably originates in the Kuiper belt objects, which orbit near the ecliptic plane in circular orbits between 30 and 50 AU, and may include Pluto as a member. The optical brightness of 2003 UB₃₁₃, if adjusted to Pluto's distance, is greater than that of Pluto, which suggested that it might be larger than Pluto². The actual size, however, could not be determined from the optical measurements because the surface reflectivity (albedo) was unknown. Here we report observations of the thermal emission of 2003 UB₃₁₃ at a wavelength of 1.2 mm, which in combination with the measured optical brightness leads to a diameter of $3,000 \pm 300 \pm 100$ km. Here the first error reflects measurement uncertainties, while the second derives from the unknown object orientation. This makes 2003 UB₃₁₃ the largest known trans-neptunian object, even larger than Pluto (2,300 km)³. The albedo is $0.60 \pm 0.10 \pm 0.05$, which is strikingly similar to that of Pluto, suggesting that the methane seen in the optical spectrum² causes a highly reflective icy surface.

The optical brightness of a planetary object derives from the reflected sunlight, and is therefore directly proportional to the square of the diameter, d , and to the optical surface reflectivity, p , which refers to the visual red and is called the 'red geometric albedo'. These three quantities are related through⁴ $d = 1,346 p^{-1/2} 10^{-H/5}$ km, where H is the absolute magnitude in the V band (around 540 nm) if the object were seen at a distance of one astronomical unit (1 AU),

which is the mean distance between Sun and Earth. The albedo is known to vary significantly between the distant minor planets, ranging from ~ 0.03 in dark trans-neptunian objects (TNOs)^{5,6} to ~ 0.6 for Pluto⁷. The optical brightness therefore provides only a lower limit to the object size when assuming $p = 1$, which in the case of 2003 UB₃₁₃ (where $H = -1.16$ mag (refs 1, 8), with a current uncertainty of about 0.1 mag) yields $d > 2,234$ km. To better constrain the size of a planetary body, it is necessary to measure its thermal emission at a wavelength where the reflected solar flux is negligible. The mm brightness is a function of the surface temperature and object size, and the temperature is only weakly dependent on the surface albedo. To measure the size of 2003 UB₃₁₃, we observed it at a wavelength of 1.2 mm, where the emission is pure thermal radiation. By combining the observed flux density at mm wavelengths with the optical brightness, one obtains a good determination of the object's diameter and geometric albedo, although some assumptions must be made for the object's unknown rotation period and orientation of the rotation axis.

Our millimetre observations were performed with the Max-Planck

Table 1 | Observation summary

Date in 2005 (day.month)	S_{ν} (250 GHz) (mJy)	τ_{zenith}	τ_{los}	t_{int} (s)
19.8	1.16 ± 0.49	0.26	0.40	6,937
23.8	0.71 ± 0.66	0.23	0.35	2,937
24.8	1.76 ± 0.54	0.23	0.34	4,273
27.8	1.27 ± 0.49	0.35	0.50	6,166
19–27.8	1.27 ± 0.26			20,313

We observed the object on four dates. Atmospheric conditions were good on 19 August (variable sky-noise), and very good on 23, 24 and 28 August. All data are used in our analysis, except for one 4-min scan during which the chopping failed. Flux densities at 250 GHz are given in mJy = 10^{-26} erg s⁻¹ cm⁻² Hz⁻¹, with a one standard deviation error giving the statistical uncertainties of the integrated measurement, which excludes systematic uncertainties due to flux calibration or pointing errors. The integration time is on-sky, half of which is on-source. The quoted zenith and line-of-sight opacities (τ_{zenith} and τ_{los} , respectively) are averaged over the integration time. The data were analysed with the MOPSIC data reduction package written by R. Zylka at IRAM. Much of the sky noise is eliminated by subtracting a weighted average signal from the neighbouring channels. The sky zenith opacity was determined through skydips, and the flux calibration was performed through frequent measurements of Uranus and Mars, which were near our target and at similar elevation. For the absolute flux calibration on the planets we find a 5% r.m.s. scatter. This calibration uncertainty must be added to the statistical errors (which decrease with integration time) of the target's flux measurement. We increase the calibration uncertainty to 10% to also account for possible pointing errors of up to 2 arcsec.

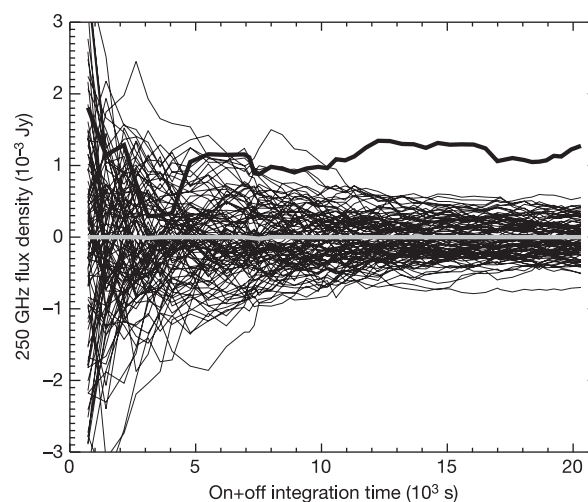


Figure 1 | Time-averaged signals of the 104 bolometers as a function of time, summarizing our 1.2 mm observations. The on-source channel is indicated by a thick black line, the median signal of the off-source channels as a thick grey line. The 117-element MAMBO-2 camera has an effective frequency for thermal radiation of 250 GHz, a half-power bandwidth of 80 GHz (210–290 GHz), and a beam size of 10.7 arcsec (corresponding to 760,000 km at a distance of 96 AU). Observations were performed using the standard on-off technique, with the sub-reflector switching every 0.25 seconds between two sky positions (on and off source) separated by 32 arcsec. The telescope pointing was frequently checked on the 5° distant quasar J0141–095 and was found to be stable within 2 arcsec.

¹Argelander Institute for Astronomy, University of Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany. ²Max-Planck-Institute for Radioastronomy, Auf dem Hügel 69, D-53121 Bonn, Germany. ³IRAM, 300 rue de la Piscine, 38406 Saint Martin d'Hères, France.

Table 2 | Properties of 2003 UB₃₁₃, Pluto, Charon and Ceres

Case	$q = A/p$	d (km)	p	T_b (K)
2003 UB ₃₁₃ , fast rotation	1	$3,136^{+263}_{-276}$	$0.54^{+0.11}_{-0.08}$	23.2
2003 UB ₃₁₃ , fast rotation	0.9	$3,094^{+268}_{-286}$	$0.55^{+0.11}_{-0.08}$	23.7
2003 UB ₃₁₃ , slow rotation	0.9	$2,859^{+231}_{-241}$	$0.65^{+0.11}_{-0.08}$	26.9
Pluto	0.9	$2,328 \pm 48$	0.62 ± 0.02	~36
Charon	0.9	$1,242 \pm 42$	0.37 ± 0.01	40
Ceres	0.4	913 ± 43	0.10 ± 0.01	164

Derived diameter, d , red geometric albedo, p , and brightness temperature, T_b , for different assumptions on rotation and phase integral, q . At mm wavelengths, fast rotation applies for objects seen equator-on and rotating with a period smaller than 40 h, whereas slow rotation applies for those seen pole-on, independent of the rotation period. Data for Pluto^{3,7,12}, Charon^{3,7,12} and Ceres¹³ are shown for comparison. Note that for Pluto the red albedo varies between ~0.49 and ~0.75 within each rotation period; for p and T_b we quote mean values.

Millimeter Bolometer (MAMBO-2) array detector at the IRAM 30 m telescope on Pico Veleta, Spain. The object was observed such that its entire emission is collected by one of the camera's horn antennae, and by switching between the object and a blank sky position twice per second. Because of the large, two-beam spacing between the array horns, this photometric (on-off) mode does measure the flux of the object, but does not provide a fully sampled image of the field around it. When averaging our flux density measurements from 19–28 August 2005 (Table 1, Fig. 1), we measure an average flux density of $S_\nu = 1.27 \pm 0.26$ mJy, where the quoted uncertainty is one standard deviation and quantifies only the averaged noise fluctuation in the signal, that is, the significance of the measurement. Systematic uncertainties must be added to this before deriving the object size and albedo. Our spatially unresolved 1.2 mm flux density observation measures the projected surface-integral of the thermal emission of the planetary body. Recent high-resolution optical observations have shown that a moon orbits 2003 UB₃₁₃, an object that is optically ~60 times fainter than the TNO itself⁹. Provided that 2003 UB₃₁₃ and its moon have a similar albedo, the moon contributes less than 2% to the observed mm flux density, a correction that is much smaller than the flux uncertainty.

Assuming that a TNO does not have a significant internal heat source, its surface-integrated thermal emission is in radiation equilibrium with the insolation. We shall further assume that such a small body cannot maintain an atmosphere that could affect the surface temperature. The temperature on the object's side facing the Sun and Earth does depend on its rotation period, the orientation of its rotation axis, and the observation wavelength. In the case of no rotation or when the rotation axis points at the Sun, the temperature on the dayside is higher than in the case of a fast rotator with equatorial view. Most known distant minor planets (TNOs, Centaurs) for which a rotation period was measured have periods of the order of hours^{5,10,11}. This justifies the 'fast rotator' approximation (period < 2 days) that the effective temperature for mm emission is independent of the 'time of day'⁵. Using Planck's radiation law and the known solar constant, the disk-averaged equilibrium temperature of a body with low heat conductivity and no atmosphere is $T_{eq} = T_0(1 - A)^{1/4}(r)^{-1/2}$, with r in units of AU; for the fast-rotator case, $T_0 = 277$ K is the disk-averaged temperature of a black body at heliocentric distance, $r = 1$ AU; for the slow rotator case, $T_0 = 329$ K. The Bond albedo, $A = q \times p$, where q is the phase integral, is a measure of the total absorbed solar energy. It is different for each Solar System object and must be determined observationally. From millimetre observations⁵ the geometric albedo, p , of TNOs is measured to range between 0.03 and 0.16, although some recent Spitzer Space Telescope observations imply even higher values⁶. The phase integral, q , relates the Bond albedo to the geometric albedo, p . Whereas $q \approx 1$ for the brightest and largest TNOs such as Pluto, the smallest TNOs can have $q \approx 0.3$ (refs 10, 11). Even though the uncertainty in q seems large, the effect on the equilibrium temperature, T_{eq} , derived from combined mm and optical measurements is small.

The flux density of the object at geocentric distance Δ is given by

$S_\nu = (\pi d^2 / 4 \Delta^2) B_\nu(T_b)$, where B_ν is Planck's law and $T_b \approx T_{eq}$ is the brightness temperature. With our observed flux density $S_\nu(250 \text{ GHz}) = 1.27 \pm 0.26$ mJy (to this statistical error we add a 10% calibration uncertainty), the relations between diameter, temperature and albedo cited above can be solved together to yield values for the diameter and albedo. The estimated 0.1 magnitude uncertainty of the optical magnitude, H , adds an additional ~1% error to the diameter and a ~7% error to the albedo. In Table 2 we list the results for different assumptions on the rotation period (or orientation) and phase integral, q . We consider the most likely case to be a fast-rotator with $q = 0.9$, resulting in a geometric albedo value very similar to that of Pluto. If we happen to view the object nearly pole-on, the slow-rotator case would apply, which then implies a smaller radius and higher albedo.

2003 UB₃₁₃ was not detected with the Spitzer Space Telescope at 70 μm wavelength, and an upper limit to the flux density was set by background source confusion (M. E. Brown, C. A. Trujillo and D. Rabinowitz, manuscript in preparation). We can predict the flux density at 70 μm given our size and albedo estimate. Because the thermal timescale of the thin surface layer responsible for the far-infrared emission is probably less than the typical rotation period of TNOs, we consider a range between the fast and slow rotation limit for computing the brightness temperature observed at 70 μm . For our favourite case $q = 0.9$, $d = 3,094$ km, $p = 0.55$, we derive a range $S_\nu(70 \mu\text{m}) = 0.7 - 2.8$ mJy between the fast and slow rotation limits, respectively, which is consistent with the observed Spitzer upper limit. In the slow rotation case, any lower value of q would imply a higher brightness temperature and higher flux density at 70 μm that would not be consistent with the observed limit. In case we actually observe the object pole-on, we predict a radius of 2,859 km and a 70 μm flux density of 1.7 mJy, which is also consistent with the Spitzer upper limit.

Received 7 October; accepted 23 November 2005.

1. Brown, M. E., Trujillo, C. A. & Rabinowitz, D. *IAU Circ.* **8577** (2005).
2. Brown, M. E., Trujillo, C. A. & Rabinowitz, D. L. Discovery of a planetary-sized object in the scattered Kuiper Belt. Preprint at (<http://arXiv.org/astro-ph/0508633>) (2005).
3. Young, E. F. & Binzel, R. Comparative mapping of Pluto's sub-Charon hemisphere — three least squares models based on mutual event lightcurves. *Icarus* **102**, 134–149 (1994).
4. Bowell, E., et al. in *Asteroids II* (eds Binzel, R., Gehrels, T. & Matthews, M.) 549–554 (Univ. Arizona Press, Tucson, 1989).
5. Altenhoff, W., Bertoldi, F. & Menten, K. M. Size estimates of some optically bright KBOs. *Astron. Astrophys.* **415**, 771–775 (2004).
6. Stansberry, J. A. et al. Albedos, diameters (and the density) of Kuiper Belt and Centaur objects. *Bull. Am. Astron. Soc.* **37**, 737 (2005).
7. Tholen, D. J. & Tedesco, E. F. Pluto's lightcurve: Results from four oppositions. *Icarus* **108**, 200–208 (1994).
8. JPL HORIZONS on-line solar system data and ephemeris computation service. (<http://ssd.jpl.nasa.gov/horizons.html>) (accessed November 2005).
9. Brown, M. E. et al. Satellites of the brightest Kuiper belt objects. Preprint at (<http://arXiv.org/astro-ph/0510029>) (2005).
10. Rabinowitz, D. L. et al. Photometric observations constraining the size, shape, and albedo of 2003 EL61, a rapidly rotating, Pluto-sized object in the Kuiper Belt. Preprint at (<http://arXiv.org/astro-ph/0509401>) (2005).
11. Sheppard, S. S. & Jewitt, D. C. Time-resolved photometry of Kuiper Belt Objects: rotations, shapes, and phase functions. *Astron. J.* **124**, 1757–1775 (2002).
12. Reinsch, K., Burwitz, V. & Festou, M. C. Albedo maps of Pluto and improved physical parameters of the Pluto-Charon system. *Icarus* **108**, 209–218 (1994).
13. Tedesco, E. F. in *Asteroids II* (eds Binzel, R., Gehrels, T. & Matthews, M.) 1090–1093 (Univ. Arizona Press, Tucson, 1989).

Acknowledgements We are grateful to the IRAM Director for providing discretionary observing time for this project, to the bolometer technology group of E. Kreysa for providing MAMBO, and to R. Zylka for the MOPIC data reduction package. We thank M. Brown for input and discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.B. (bertoldi@astro.uni-bonn.de).

A low density of 0.8 g cm^{-3} for the Trojan binary asteroid 617 Patroclus

Franck Marchis¹, Daniel Hestroffer², Pascal Descamps², Jérôme Berthier², Antonin H. Bouchez³, Randall D. Campbell³, Jason C. Y. Chin³, Marcos A. van Dam³, Scott K. Hartman³, Erik M. Johansson³, Robert E. Lafon³, David Le Mignant³, Imke de Pater¹, Paul J. Stomski³, Doug M. Summers³, Frédéric Vachier², Peter L. Wizinovich³ & Michael H. Wong¹

The Trojan population consists of two swarms of asteroids following the same orbit as Jupiter and located at the L4 and L5 stable Lagrange points of the Jupiter–Sun system (leading and following Jupiter by 60°). The asteroid 617 Patroclus is the only known binary Trojan¹. The orbit of this double system was hitherto unknown. Here we report that the components, separated by 680 km, move around the system's centre of mass, describing a roughly circular orbit. Using this orbital information, combined with thermal measurements to estimate the size of the components, we derive a very low density of $0.8^{+0.2}_{-0.1} \text{ g cm}^{-3}$. The components of 617 Patroclus are therefore very porous or composed mostly of water ice, suggesting that they could have been formed in the outer part of the Solar System².

The existence of binary asteroidal systems has been confirmed observationally during the past decade with spacecraft exploration³, ground-based imaging⁴, radar observations⁴, and light curve measurements⁵. Most recently, the discovery of two moons orbiting around the irregular rubble pile asteroid 87 Sylvia⁶ with adaptive optics system observations confirms that collision and disruption is the main formation mechanism for multiple main-belt systems. The system 617 Patroclus, the only binary Trojan known¹, was discovered with the Hokupa'a Gemini 8-m adaptive optics system⁷ under excellent seeing conditions. Because of their faintness (with a magnitude in the visible spectrum of $m_v > 15.5$), Trojan asteroids cannot be directly observed by most of the adaptive optics systems. Using related techniques such as stellar appulse⁸ and Laser Guide Star (LGS) observations⁹ at the Lick 3-m telescope, our group failed to detect any new companions of Trojans, indicating that the proportion of multiple systems in the Trojan population larger than 40 km in diameter is less than 4% (ref. 9). The study of a companion's orbit provides information about the mass, density and porosity of a system, illuminating the role of these small bodies in the formation of our Solar System.

In 2004, a LGS¹⁰ adaptive optics system was offered on the Keck II, a 10-m telescope at the summit of Mauna Kea in Hawaii, USA. Thanks to this technology, adaptive optics sky coverage has been expanded and a Trojan asteroid, such as 617 Patroclus ($m_v \approx 16$) can be observed at any time under correct seeing conditions and visibility. We initiated an observing campaign between November 2004 and May 2005, recording direct images of the double system in H (1.6 μm) and Kp (2.2 μm) broadband filters with the NIRC2 near-infrared camera (Fig. 1). Both components of the system were detected at five different epochs with an angular separation of between 45 and 190 milli-arcsecond (mas) and a difference in

magnitude of $\Delta m \approx 0.17$. Supplementary Table 1 summarizes the relative positions and brightnesses of the components. To refine our model and check its consistency, we included four additional observations from the Gemini Science archive taken in 2001¹ and 2002 with Hokupa'a⁷ mounted on Gemini North. In excellent seeing conditions, this natural guide star low-order adaptive optics system provides an angular resolution of only 85 mas.

The orbital parameters of the system were estimated independently using two algorithms for visual binaries, a Monte Carlo technique¹¹ (Fig. 2) and a generalized least-squares method¹². Both methods have been successfully applied to several main-belt binary systems¹³, such as 121 Hermione¹⁴ and the moonlets of 87 Sylvia⁶. The solution with the orbital elements indicated in Table 1 is quite

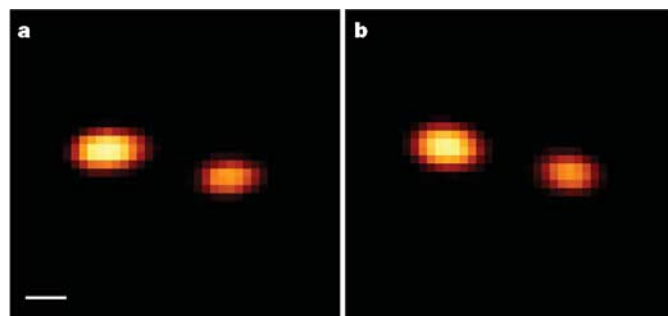


Figure 1 | 617 Patroclus observed with the Keck 10-m telescope on the Mauna Kea summit in Hawaii and its adaptive optics system. The observations of this faint object (integrated magnitude in the visible, $m_v \approx 15.8$) were made possible using the LGS system. An artificial star (with $m_v \approx 11$) is created at proximity of the target, exciting the thin sodium layer at an altitude of ~ 100 km using a dye laser with a power output of ~ 13 W. The asteroid itself was used for the tip-tilt analysis, which is acquired by an avalanche photodiode tip-tilt sensor, and then corrected by balancing a flat mirror. The lowest-order aberrations are corrected by controlling a 349-actuator deformable mirror at a rate of 400 Hz. These observations were taken on 28 May 2005 with the NIRC-2 infrared camera through the H, centred at 1.6 μm (a), and Kp, centred at 2.2 μm (b), broadband filters. The angular resolution of the data is estimated to 58 mas, close to the diffraction limit of the telescope (indicated by the scale bar, 50 mas). In these two images, the angular separation of both components is 150 mas, corresponding to 640 km, close to the maximum separation of the system. The magnitude difference in brightness is 0.2 in both filters, suggesting a similar surface composition.

¹Department of Astronomy, University of California, 601 Campbell Hall, Berkeley, California 94720, USA. ²Institut de Mécanique Céleste et de Calculs des Éphémérides, UMR CNRS 8028, Observatoire de Paris, 77 Avenue Denfert-Rochereau, F-75014 Paris, France. ³W. M. Keck Observatory, 65-1120 Mamalahoa Highway, Kamuela, Hawaii 96743, USA.

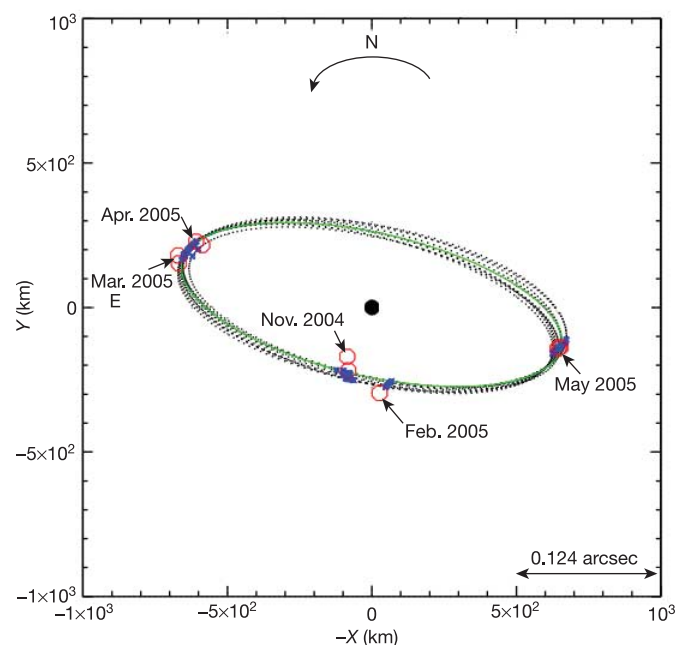


Figure 2 | Several good-fit orbits of the 617 Patroclus components estimated using 2004–2005 Keck LGS adaptive optics data. No correction for parallax was made. The least-squares orbit (green) corresponds to the best-fit solution that minimizes the root-mean-square residual. The orbital parameters of the orbit were validated and refined by adding the Gemini/Hokupa'a observations taken in 2001 and 2002. The period is 4.283 days.

accurate considering 2004–2005 data with a root-mean-square residual error of 9 mas, corresponding to ~ 35 km. It is slightly degraded when the lower-angular-resolution 2001–2002 data provided by our colleagues¹ and from the Gemini North telescope archive are included (20 mas).

The two components, separated by 680 ± 20 km, revolve around their centre of mass in 4.283 ± 0.004 days in a roughly circular orbit ($e \approx 0.02 \pm 0.02$). Because we could obtain a purely keplerian solution matching four years of observations, we can deduce that the pole orientation of the orbit seems not to precess, and is thus probably aligned with the angular momentum pole. The low eccentricity of the orbit, compared for instance with the extremely high eccentricity (~ 0.8) of the binary Kuiper Belt Object¹⁵ 1998WW31, indicates that dissipation effects, such as tides, must be considered to circularize the orbit of this double Trojan system. The angular resolution provided by the Keck LGS adaptive optics data (0.06 arcsec) is above the angular size of the components. Their size can be estimated only using near-simultaneous mid-infrared and visible observations. When the 617 Patroclus system was observed in November 2000¹⁶, our orbital model shows that both components were clearly separated, meaning that the estimated radiometric radius (R) corresponds to the sum of the radii of the components through the relation $R^2 = R_1^2 + R_2^2$. Assuming the same albedo for the components (a realistic assumption because the brightness ratio in H and K bands are constant and the objects are spherical) and a size ratio of $R_1/R_2 = 1.082$, we derive a radius of $R_1 = 60.9$ km and $R_2 = 56.3$ km (with an error of 1.6 km and an albedo in the visible spectrum of $A_v = 0.04$) and a beaming parameter $\eta = 0.94$. Recent Spitzer spectra taken between 5 and $40 \mu\text{m}$ of Trojan asteroids¹⁷ confirm that the effect of the beaming factor is negligible ($\eta \approx 1$).

Using the total mass ($M = 1.36 \times 10^{18}$ kg) derived from our orbital parameters and the size of the components estimated by radiometric measurements, we can derive for the first time the average bulk density of a Trojan P-type asteroid, a class of object whose interior composition is generally not well known but may

Table 1 | Orbital elements of the binary system 617 Patroclus

Period (days)	4.283 ± 0.004
Semi-major axis (km)	680 ± 20
Eccentricity, e	0.02 ± 0.02
Orbit pole solution ECJ2000 (λ and β in degrees)	$\lambda = 234 \pm 5$ $\beta = -62 \pm 1$
Pericentre argument (degrees)	139
Time of passage (Julian days, JD)	2453171.1 ± 0.1
Total mass (kg)	$(1.36 \times 10^{18}) \pm 0.11$
Bulk density of the system* (g cm^{-3})	$0.8^{+0.2}_{-0.1}$

Two models^{11,12} were used to estimate the orbital parameters (with 1σ error bars) using 2004–2005 Keck LGS adaptive optics data and give a consistent solution. The addition of 2001 and 2002 observations taken with the Gemini North 8-m telescope and the Hokupa'a using our general least-squares method¹² improves by a factor of ten the accuracy on the period, removes the ambiguity on the pole solution of the orbit, which is retrograde, and confirms the absence of measurable precession, suggesting a low inclination of the system with respect to its invariant plane.

*Considering the radiometric measurements¹⁶ with a beaming factor $\eta = 0.89$ –0.96 and assuming the shape of the two components to be spherical with $R_1 = 60.9$ km and $R_2 = 56.3$ km.

possibly include organic-rich silicates, carbon and anhydrous silicates, and water ice. Even considering the uncertainty in the volume of the components which accounts for most of the error bar, the density of 617 Patroclus ($\rho = 0.8^{+0.2}_{-0.1} \text{ g cm}^{-3}$) is extremely low, if compared with the bulk-density of known binary C-type main-belt asteroids¹³ with $\rho \approx 1.2 \text{ g cm}^{-3}$ such as 45 Eugenia^{18,19}, 90 Antiope^{19,20}, 87 Sylvia⁶, or 121 Hermione¹⁴. The knowledge of the mean density places a strong constraint on the internal structure and composition of this Trojan. The Trojan swarms are currently at a distance from the Sun similar to that of the Galilean satellites, so they may be made of the same material (a mixture of rock and ice such as in Ganymede and Callisto) with an uncompressed sample density²¹ of $\sim 1.6 \text{ g cm}^{-3}$. This yields a macro-porosity of $\sim 50\%$ (see Fig. 3), typical of loose-rubble-pile main-belt asteroids²². A more realistic smaller porosity of $\sim 15\%$ cannot be excluded by considering a different composition, such as more water ice, in the interior of 617 Patroclus. Dynamical simulations² show that Trojans could have formed in more distant regions and then been captured into co-orbital motion with Jupiter during the time when the giant planets migrated by interaction with a surrounding disk of planetesimals. Quantitatively, considering the observed size distribution of the Trojan population and the present bulk-density measurement of 617 Patroclus, the total mass of the Trojans calculated ($7 \times 10^{-6} M_{\text{Earth}}$) is also into the range of the total mass of their simulation² ($4 \times 10^{-6} M_{\text{Earth}}$ to $3 \times 10^{-5} M_{\text{Earth}}$).

An additional feature that cannot be determined by this study is the obliquity of the components' equator with respect to the determined orbital plane. It is therefore not possible to see whether the system reached its Cassini state²³ (a resonance between spin precession and orbital precession), or whether the spins are prograde or retrograde with respect to the orbit, and with none or a significant obliquity. Because no complete light curve variations have been successfully recorded, the spin periods of the components remain unknown. Photometric observations²⁴ suggest two non-commensurate periods, implying an unsynchronized system. Because it is still unclear whether the system is synchronous, like 90 Antiope^{20,25} (the only known resolved double main-belt asteroid), a special programme of photometric observations should be undertaken to attempt to derive the spin poles and rates.

The orbital angular momentum of this binary system, which would account for more than 90% of the total angular momentum if we assume a synchronous orbit, is close to $(2.4 \text{ G m}^3 \text{ s}^{-1})^{1/2}$, which is well in excess of the limit of $J_{\text{cr}} = 0.4$ for binary fission of a rotating mass of fluid²⁶, and is also larger than that of the other known double asteroid 90 Antiope²⁷. The origin of this tightly bound (the separation corresponds to $\sim 1/60 \times R_{\text{Hill}}$), small-eccentricity and large-angular-momentum binary system remains unclear. It could be the result of a capture as proposed in the case of binary minor planets in the Kuiper belt^{28,29}, with subsequent circularization of the orbit by

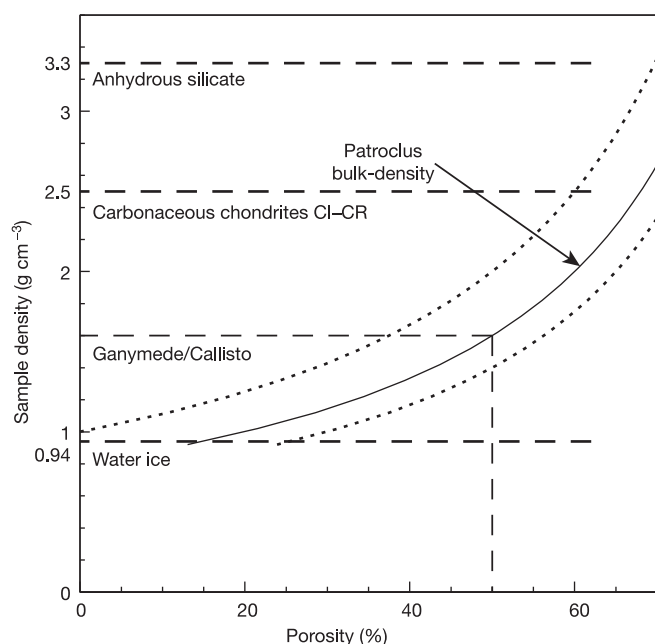


Figure 3 | Relationship between the sample density and the porosity of 617 Patroclus for various compositions. Using the orbital information of the two components, combined with the thermal measurements¹⁶, we derived a very low bulk density of $\sim 0.8 \text{ g cm}^{-3}$ (solid line). The dotted lines correspond to the uncertainty in the bulk density for a plausible range of beaming factors varying from 0.89 to 0.96, compatible with Spitzer Trojan spectra¹⁷. If 617 Patroclus components were made of the same ice/rock mixture as Ganymede or Callisto²¹, it would be extremely porous (porosity $\approx 50\%$). If they were composed of pure water ice, its porosity would be more realistic (porosity $\approx 15\%$) but would imply its formation in a more distant part of the Solar System².

tidal effects. However, a low-velocity impact of a relatively large projectile could also have supplied enough angular momentum to produce a fission of the parent body without completely dispersing it. The relative velocities in the Trojan swarms are, however, at present several kilometres per second, making such a scenario less probable. Finally, this system shares some similarities with known binary near-Earth asteroids (having small separation, mass ratios close to unity, and circular orbits), implying that it could have also formed by tidal splitting after a low-velocity encounter with a large planet³⁰. Our colleagues observed in their simulations² that several distant planetesimals have a close approach with Jupiter before their final capture at the Lagrange points.

Received 19 September; accepted 17 October 2005.

- Merline, W. J. *et al.* S/2001 (617) 1. *IAU Circ.* **7741** (2001).
- Morbidelli, A., Levison, H. F., Tsiganis, K. & Gomes, R. Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* **435**, 462–465 (2005).
- Chapman, C. R. *et al.* Discovery and physical properties of Dactyl, a satellite of asteroid 243 Ida. *Nature* **374**, 783–785 (1995).
- Merline, W. J. L., *et al.* in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 289–312 (Univ. Arizona Press, Tucson, 2002).
- Pravec, P. *et al.* Photometric survey of binary near-Earth asteroids. *Am. Astron. Soc. Div. Planet. Sci.* **36**, abstr. 28.04 (2004).
- Marchis, F., Descamps, P., Hestroffer, D. & Berthier, J. Discovery of the triple asteroidal system 87 Sylvia. *Nature* **436**, 822–824 (2005).
- Roth, K. C. *et al.* Hokupa'a performance and point spread function characterization. *Bull. Am. Astron. Soc.* **33**, abstr. 02.02, 785 (2001).

- Berthier, J. & Marchis, F. A tool for observations of Centaurs/Kuiper Objects with adaptive optics systems. *Bull. Am. Astron. Soc.* **33**, abstr. 12.17, 1049 (2001).
- Marchis, F. *et al.* in *Advancements in Adaptive Optics* (eds Bonaccini, D., Ellerbroek, B. L. & Ragazzoni, R.) *Proc. SPIE* **5490**, 338–350 (2004).
- Bouchez, A. H. *et al.* in *Advancements in Adaptive Optics* (eds Bonaccini, D., Ellerbroek, B. L. & Ragazzoni, R.) *Proc. SPIE* **5490**, 321–330 (2004).
- Hestroffer, D. & Vachier, F. Orbit determination of binary asteroids. *IAU Symp.* **229 on Asteroids, Comets and Meteors (Buzios, Rio de Janeiro, Brazil, 2005)** abstr. 10.9, 87 (IAU, 2005).
- Descamps, P. Orbit of an astrometric binary system. *Celest. Mech. Dynam. Astron.* **92**, 381–402 (2005).
- Marchis, F., *et al.* On the diversity of binary asteroid orbits. *IAU Symp.* **229 on Asteroids, Comets and Meteors (Buzios, Rio de Janeiro, Brazil, 2005)** abstr. 10.1, 83 (IAU, 2005).
- Marchis, F. *et al.* Mass and density of asteroid 121 Hermione from an analysis of its companion orbit. *Icarus* (in the press).
- Veillet, C. *et al.* The binary Kuiper-belt 1998 WW31. *Nature* **416**, 711–713 (2000).
- Fernandez, Y. *et al.* The albedo distribution of jovian trojan asteroids. *Astron. J.* **126**, 1563–1574 (2003).
- Emery, J. P., Cruikshank, D. P. & Van Cleve, J. Thermal emission spectroscopy of asteroids with the Spitzer Space Telescope. *Am. Astron. Soc. Div. Planet. Sci.* **37**, abstr. 15.07 (2005).
- Merline, W. J. L. *et al.* Discovery of a moon orbiting the asteroid 45 Eugenia. *Nature* **401**, 565–569 (1999).
- Marchis, F., Descamps, P., Berthier, J., Hestroffer, D. & de Pater, I. Fine analysis of 121 Hermione, 45 Eugenia, and 90 Antiope binary asteroid systems with AO observations. *Am. Astron. Soc. Div. Planet. Sci.* **36**, abstr. 46.02 (2004).
- Descamps, P., *et al.* Insights on 90 Antiope double asteroid combining VLT-AO and lightcurve observations. *IAU Symp.* **229 on Asteroids, Comets and Meteors (Buzios, Rio de Janeiro, Brazil, 2005)** abstr. 10.7, 87 (IAU, 2005).
- Anderson, J. D. *et al.* Amalthea's density is less than that of water. *Science* **308** (5726), 1291–1293 (2005).
- Britt, D. T., Yeomans, D., Housen, K. & Consolmagno, G. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 485–500 (Univ. Arizona Press, Tucson, 2002).
- Gladman, B., Quinn, D., Nicholson, P. & Rand, R. Synchronous locking of tidally evolving satellites. *Icarus* **122**, 166–172 (1996).
- Angeli, C. A. *et al.* A contribution to the study of asteroids with long rotation period. *Planet. Space Sci.* **47**, 699–714 (1999).
- Michalowski, T. *et al.* Eclipsing binary asteroid 90 Antiope. *Astron. Astrophys.* **423**, 1159–1168 (2004).
- Tohline, J. E. The origin of binary stars. *Annu. Rev. Astron. Astrophys.* **40**, 349–385 (2002).
- Weidenschilling, S. J., Marzari, F., Davis, D. R. & Neese, C. Origin of the double asteroid 90 Antiope: a continuing puzzle. *Lunar Planet. Sci.* **XXXII**, 1890 (2001).
- Goldreich, P., Lithwick, Y. & Re'em, S. Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* **410**, 643–646 (2002).
- Astakhov, S. A., Lee, E. A. & Farrelly, D. Formation of Kuiper-belt binaries through multiple chaotic scattering encounters with low-mass intruders. *Mon. Not. R. Astron. Soc.* **360**, 401–415 (2005).
- Walsh, K. J. & Richardson, D. C. Binary near-Earth asteroid formation: rubble pile model of tidal disruptions. *Am. Astron. Soc. Div. Planet. Sci.* **37**, abstr. 14.11 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Science Foundation Science and Technology Center for Adaptive Optics and by the National Aeronautics and Space Administration (NASA) issue through the Science Mission Directorate Research and Analysis programmes. Most of the data were obtained at the W. M. Keck observatory, which is operated as a scientific partnership between the California Institute of Technology, the University of California and NASA. Additional observations were obtained at the Gemini Observatory (acquired through the Gemini Science Archive).

Author Contributions F.M. and the IMCCE researchers processed, analysed and interpreted the data. The 2004–2005 campaign of observations with Keck LGS adaptive optics was conducted by the team from the W. M. Keck Observatory, and other University of California at Berkeley researchers.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.M. (e-mail: fmarchis@berkeley.edu).

LETTERS

Disruption of extended defects in solid oxide fuel cell anodes for methane oxidation

Juan Carlos Ruiz-Morales^{1†}, Jesús Canales-Vázquez^{1†}, Cristian Savaniu¹, David Marrero-López², Wuzong Zhou¹ & John T. S. Irvine¹

Point defects largely govern the electrochemical properties of oxides: at low defect concentrations, conductivity increases with concentration; however, at higher concentrations, defect-defect interactions start to dominate^{1,2}. Thus, in searching for electrochemically active materials for fuel cell anodes, high defect concentration is generally avoided. Here we describe an oxide anode formed from lanthanum-substituted strontium titanate (La-SrTiO_3) in which we control the oxygen stoichiometry in order to break down the extended defect intergrowth regions and create phases with considerable disordered oxygen defects. We substitute Ti in these phases with Ga and Mn to induce redox activity and allow more flexible coordination. The material demonstrates impressive fuel cell performance using wet hydrogen at 950 °C. It is also important for fuel cell technology to achieve efficient electrode operation with different hydrocarbon fuels^{3,4}, although such fuels are more demanding than pure hydrogen. The best anode materials to date—Ni-YSZ (yttria-stabilized zirconia) cermet⁵—suffer some disadvantages related to low tolerance to sulphur⁶, carbon build-up when using hydrocarbon fuels⁷ (though device modifications and lower temperature operation can avoid this^{8,9}) and volume instability on redox cycling. Our anode material is very active for methane oxidation at high temperatures, with open circuit voltages in excess of 1.2 V. The materials design concept that we use here could lead to devices that enable more-efficient energy extraction from fossil fuels and carbon-neutral fuels.

Over the past few years there has been a growing interest in perovskite-based materials in the search for alternative materials to Ni-YSZ cermets as fuel electrodes in solid oxide fuel cells (SOFCs)^{3,10}. Such perovskites are normally oxygen stoichiometric or substoichiometric: here we seek to explore perovskites that are nominally oxygen overstoichiometric as SOFC fuel electrodes. Initially we focus on phases with extended oxygen-rich defects, and then attempt to destabilize these extended defects by reducing the degree of oxygen excess. Titanates with nominal oxygen overstoichiometry are especially interesting owing to their very high electronic conductivity, stability in reducing conditions and resistance to sulphur poisoning^{11–14}. Oxides belonging to the $\text{La}_{1-x}\text{Sr}_x\text{TiO}_{3+\delta}$ system have been previously studied as anode materials showing moderate performance compared to those of the state of the art materials, but better catalytic properties and ionic conductivity are desirable^{11,12}.

The performance of titanate-based materials was greatly improved by using composites with CeO_2 (ref. 15), based on the catalytic properties of the ceria. These lanthanum strontium titanates are usually treated in the literature as simple cubic perovskites, although the presence of extra oxygen beyond the ABO_3 stoichiometry plays a

critical role in both the structure and the electrochemical properties, as summarized in Fig. 1. The lower members of the $\text{La}_4\text{Sr}_{n-4}\text{Ti}_n\text{O}_{3n+2}$ series, $n < 7$, are layered phases, having oxygen-rich planes in the form of crystallographic shears joining consecutive

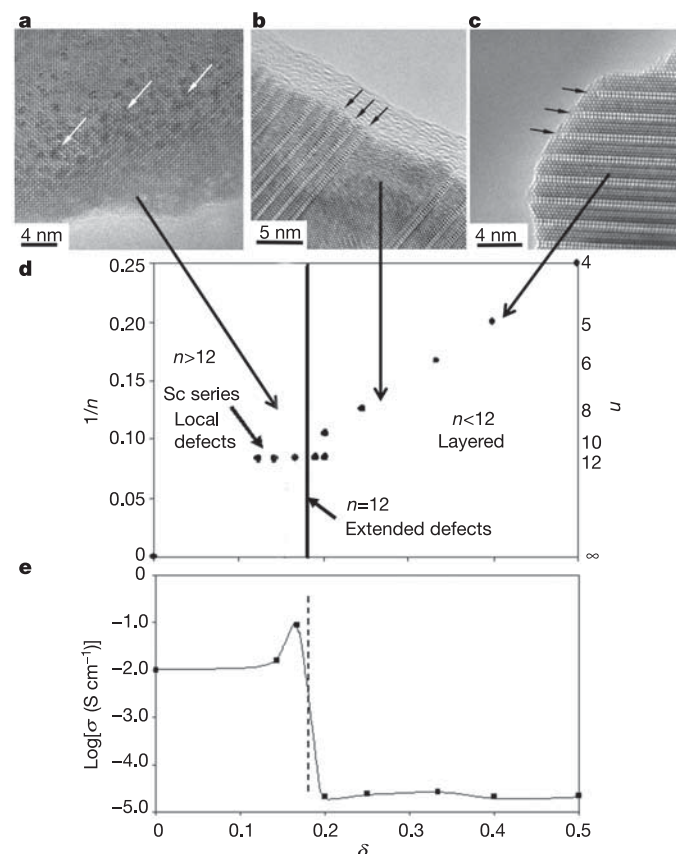


Figure 1 | Relation between microstructure, composition and conductivity of the ' $\text{La}_4\text{Sr}_{n-4}\text{Ti}_n\text{O}_{3n+2}$ ' series. **a–c**, HRTEM images of samples varying from disordered extended defects (**a**, $n = 12$) through random layers of extended defects (**b**, $n = 8$) to ordered extended planar oxygen excess defects (**c**, $n = 5$). **d**, The location of these phases on the composition map, with $1/n$ plotted against oxygen excess δ in perovskite unit $\text{ABO}_{3+\delta}$. **e**, Defect electronic conductivity of grain component as determined by a.c. impedance spectroscopy on samples quenched from 1,300 °C in air, also plotted against δ . 'Sc series' refers to samples in series $(\text{La}_4\text{Sr}_{n-4})(\text{Ti}_{n-y}\text{Sc}_y)\text{O}_{3n+2-y/2}$ (ref. 14).

¹School of Chemistry, University of St Andrews, St Andrews, Fife KY16 9ST, UK. ²Department of Inorganic Chemistry, University of La Laguna, 38200, Canary Islands, Spain.

[†]Present addresses: Department of Inorganic Chemistry, University of La Laguna, 38200, Canary Islands, Spain (J.C.R.-M.); Instituto de Ciencia de los Materiales ICMA-B-CSIC, 08193, Barcelona, Spain (J.C.-V.).

Table 1 | Properties of $\text{La}_4\text{Sr}_8\text{Ti}_{12-x}\text{M}_x\text{O}_{38-6}$ electrodes

Electrode	R_p ($\Omega \text{ cm}^2$)		OCV (V)	Representative electrode grain size (μm)
	97.7% H_2 /2.3% H_2O	97.7% CH_4 /2.3% H_2O		
$\text{La}_4\text{Sr}_8\text{Ti}_{12}\text{O}_{38-6}$, $x = 0^*$	2.97	8.93	0.98	1
$\text{M} = \text{Sc}$, ($x = 0.6$) †	0.50	1.20	1.10	0.5–1.0
$\text{M} = \text{Mn}$ ($x = 1$) ‡	0.43	1.14	0.98	0.8
Mn ($x = 0.5$), Ga ($x = 0.5$)	0.20	0.57	1.25	0.8

Comparison of open circuit voltage (OCV) and polarization resistance (R_p) measured in indicated gas compositions at 900 °C on doped and undoped $\text{La}_4\text{Sr}_{n-4}\text{Ti}_n\text{O}_{3n+2}$.

*Ref. 11; † ref. 14; ‡ ref. 23.

blocks. These planes become more sporadic with increasing n (that is, decreasing the oxygen content) until they are no longer crystallographic features, rendering local oxygen-rich defects randomly distributed within a perovskite framework, $n > 11$ (ref. 16). Substitution of Ti^{4+} by Nb^{5+} or Sc^{3+} demonstrates that the oxygen excess parameter (δ) critically determines whether defects are ordered or disordered, with $\delta = 0.167$ being a critical parameter. The presence of such disordered defects appears to strongly affect the redox characteristics of the oxide, as indicated by marked effects on conductivity induced by mild reduction (Fig. 1). Although the materials from this lanthanum strontium titanate oxygen excess series are much easier to reduce, and hence exhibit much higher electronic conductivity than their oxygen stoichiometric analogues, they do not exhibit electrochemical performance comparable to such an effective oxide anode as $\text{La}_{0.75}\text{Sr}_{0.25}\text{Cr}_{0.5}\text{Mn}_{0.5}\text{O}_{3-6}$ (ref. 10). This we attribute to the inflexibility of the co-ordination demands of titanium, which strongly prefers octahedral coordination in the perovskite environment.

In order to make the B-site co-ordination more flexible and hence to improve electrocatalytic performance, we have introduced various dopants to replace Ti in $\text{La}_4\text{Sr}_8\text{Ti}_{12}\text{O}_{38-6}$ based fuel electrodes, Table 1, and found the most successful to be a combination of Mn and the trivalent Ga ion. Mn supports p-type conduction in oxidizing conditions, and has been previously shown to promote electroreduction under SOFC conditions^{17,18}. Furthermore, Mn is known to accept lower coordination numbers in perovskites¹⁹,

especially for Mn^{3+} , and thus it may facilitate oxide-ion migration. Similarly Ga is well known to adopt lower co-ordination than octahedral in perovskite-related oxides. The possibility of mixed ionic/electronic conduction is very important, because it would allow the electro-oxidation process to move away from the three-phase electrode/electrolyte/gas interface onto the anode surface, with considerable catalytic enhancement³.

$\text{La}_4\text{Sr}_8\text{Ti}_{11}\text{Mn}_{0.5}\text{Ga}_{0.5}\text{O}_{37.5}$ (LSTMG) powders were prepared by solid state reaction from stoichiometric amounts of high purity La_2O_3 , SrCO_3 , TiO_2 , Mn_2O_3 and Ga_2O_3 fired for 24–48 h. Polarization measurements were performed in a three-electrode arrangement. The electrolyte was a sintered 8 mol% Y_2O_3 stabilized ZrO_2 (YSZ) pellet of 2 mm thickness and 20 mm diameter. $\text{La}_{0.8}\text{Sr}_{0.2}\text{MnO}_3$ (LSM) and Pt were used as counter-electrode and reference electrode, respectively. The anode was prepared in two configurations, first as a 60- μm -thick layer of 50:50 LSTMG:YSZ and second as an optimized anode with four layers, with graded concentration of YSZ. Each layer was pre-fired at 300 °C and all of them co-fired at 1,200 °C for 2 h. Electrical contacts to the anodes were formed using an Au mesh with small amounts of Au paste to ensure contacting and to avoid any additional catalytic effect. Fuel-cell performance was obtained for these materials as SOFC anodes with 330- μm -thick YSZ(Al_2O_3) electrolyte and LSM cathode. Platinum paste coated onto LSM and fired at 900 °C was used as the cathode current collector.

$\text{La}_4\text{Sr}_8\text{Ti}_{11}\text{Mn}_{0.5}\text{Ga}_{0.5}\text{O}_{37.5}$ forms as a single-phase perovskite on firing at 1,400 °C. The structure obtained can be refined as monoclinic, with $a = 5.5287(9)$ Å, $b = 7.8098(5)$ Å, $c = 5.3096(6)$ Å, $\beta = 92.295(7)^\circ$, $V = 229.08(6)$ Å³. No chemical reactions were observed by X-ray diffraction on firing an intimate mixture of LSTMG and YSZ pressed powder at 1,200 °C in air for 80 h, indicating good chemical compatibility. The phase is stable under fuel conditions at 1,000 °C; the perovskite structure is retained on

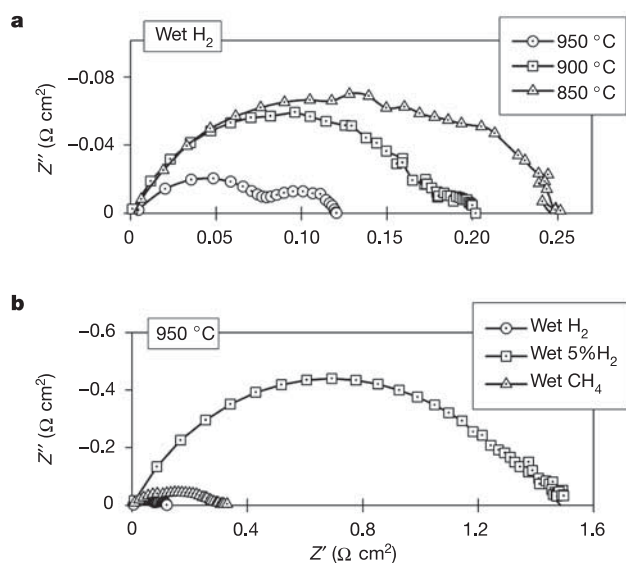


Figure 2 | Polarization measurements on LSTMG/YSZ with varying temperatures and atmospheres. Polarization impedances were measured **a**, under humidified hydrogen at different temperatures, and **b**, under different atmospheres at 950 °C, all humidified at 20 °C on a screen-printed graded $\text{La}_4\text{Sr}_8\text{Ti}_{11}\text{Mn}_{0.5}\text{Ga}_{0.5}\text{O}_{37.5}/\text{YSZ}$ working electrode on a 2 mm YSZ electrolyte with $\text{La}_{0.8}\text{Sr}_{0.2}\text{MnO}_3$ counter and reference electrodes. Z' and Z'' are respectively the real and the imaginary parts of the complex impedance.

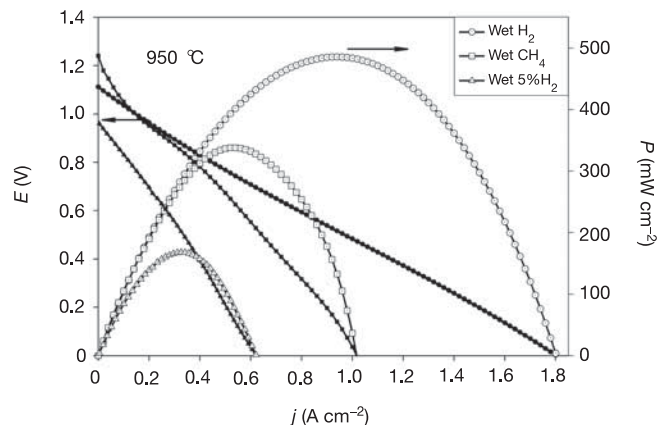


Figure 3 | Performance plots in different atmospheres. Fuel cell performance plots for different fuel gas compositions each containing 2.3% of water at a four-layer optimized LSTMG anode (circle: pure H_2 , square: pure CH_4 , triangle: 5% H_2) on a 330 μm thin YSZ electrolyte, with LSM cathode in unhumidified oxygen. E is potential difference between electrodes, j is current density and P is power density.

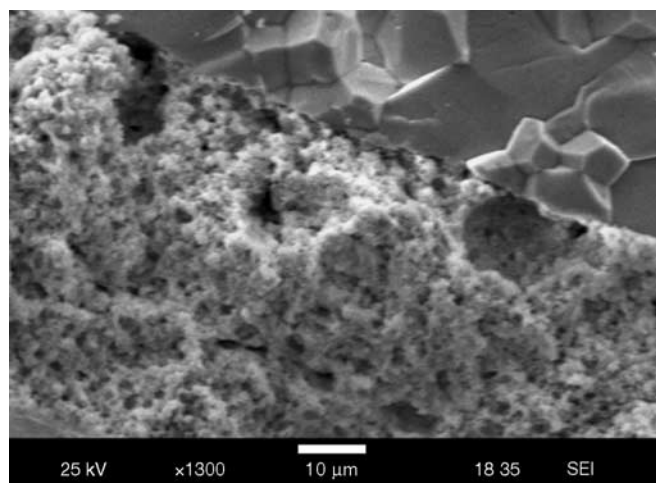


Figure 4 | Electrode interface. Scanning electron microscope image, showing the cross-section of a fuel cell after testing.

firing in 4.9% H_2 /2.3% H_2O /92.8%Ar (thereafter termed wet 5% H_2) for 24 h, with the cell changing to orthorhombic $a = 5.5343(6) \text{ \AA}$, $b = 7.8239(4) \text{ \AA}$, $c = 5.5322(7) \text{ \AA}$, $V = 239.54(6) \text{ \AA}^3$. Detailed high-resolution transmission electron microscopy (HRTEM) studies show exactly the same features that were observed for $\text{La}_4\text{Sr}_8\text{Ti}_{12}\text{O}_{38}$ ($n = 12$), that is, this phase is within the localized defect region, as predicted in Fig. 1 for $\delta < 0.167$. On re-oxidation of LSTMG in air, in a thermogravimetric analyser, a weight increase of 0.37% was observed in a step between 400 and 600 °C, corresponding to a change in oxygen content of about 0.5 oxygen atoms per formula unit, consistent with Mn^{II} to Mn^{IV} reoxidation. Some further weight gain of about 0.07% (0.1 O) was observed on holding at 900 °C which may be related to Ti^{III} to Ti^{IV} reoxidation and is fairly typical for these titanates¹⁴.

The total conductivity is 0.004 S cm^{-1} in air at 900 °C and increases up to 0.5 S cm^{-1} at 10^{-15} atm . Activation energies for conduction are 0.16 eV for 200–700 °C, and 0.56 eV for 700–900 °C in wet 5% H_2 . The conductivity of $\text{La}_4\text{Sr}_8\text{Ti}_{11}\text{Mn}_{0.5}\text{Ga}_{0.5}\text{O}_{37.5}$ was also measured as a function of oxygen partial pressure at 800 °C and 900 °C. The material shows typical n-type conductivity as the dominant conduction mechanism, but some evidence of p-type conductivity can be found at $p_{\text{O}_2} > 0.5 \text{ atm}$.

The anode polarization resistance was measured using three-electrode geometry as previously described¹¹. Results at different temperatures in humidified hydrogen are shown in Fig. 2a and in different atmospheres at 950 °C in Fig. 2b using an optimized anode made of four graded layers with varying LSTMG:YSZ ratios and total thickness 140 μm . The polarization resistances in wet H_2 (97.7% H_2 , 2.3% H_2O) were $0.12 \Omega \text{ cm}^2$ at 950 °C, $0.20 \Omega \text{ cm}^2$ at 900 °C and $0.25 \Omega \text{ cm}^2$ at 850 °C. The polarization resistances were $0.12 \Omega \text{ cm}^2$ in wet H_2 , $1.5 \Omega \text{ cm}^2$ in wet 5% H_2 and a remarkably low value, $0.36 \Omega \text{ cm}^2$, in wet CH_4 (97.7% CH_4 , 2.3% H_2O), at 950 °C. These polarization resistances were attained after about 24 h in fuel conditions; initial polarization resistances were 2–3 times higher. This long time period to achieve equilibration is fairly typical for donor doped strontium titanates that are not cation vacancy compensated, and we attribute this to reorganization of a complex defect structure. The best previous results using metal-free oxide anodes only achieved polarization resistances of twice these values and with much lower OCVs (open circuit voltages) in methane³. These polarization resistances are about 15 times less than previously achieved without Mn/Ga addition¹¹, and less than 50% of the values obtained using just Mn, all for electrodes with similar microstructures (Table 1). Furthermore, these results are comparable with the best cermet electrode performances, and allow both operation in low steam hydrocarbon fuels and good mechanical redox stability.

The OCVs matched the value predicted by the Nernst equation, 0.97 V and 1.13 V at 950 °C, for wet 5% H_2 and wet H_2 , respectively. The OCVs in wet CH_4 , for a single layer 50:50 YSZ:LSTMG anode, were: 1.39 V at 950 °C, 1.32 V at 900 °C and 1.36 V at 850 °C. These values were reproducible after two days of testing in wet 5% H_2 , wet H_2 and wet CH_4 . For a four-layer anode, in wet CH_4 , the OCVs were: 1.23 V, 1.17 V and 1.16 V at 950 °C, 900 °C and 850 °C, respectively. The high OCVs compared to those typically observed with methane fuels are more significant than, but broadly similar to, that for the addition of precious metal to copper ceria anodes²⁰. It seems clear that the thicker, more complex anode structure is less able to fully activate methane at the electrode/electrolyte interface, even though the electrochemical performance of the complex anode is superior in terms of polarization resistance. The OCV trend observed at the thicker electrode is similar to that obtained by Liu and Barnett²¹ which they initially attributed to partial oxidation of produced carbon, but later²² suggested was due to the oxidation of hydrogen, produced by methane reforming with the humidified methane reaching equilibrium at the anode. Here the high OCV we obtained at the thin electrode (1.4 V) implies full oxidation of equilibrated wet CH_4 , as this potential is the expected OCV in such an atmosphere.

Figure 3 shows the performance of the LSTMG anode in different atmospheres, at 950 °C, using a two-electrode set-up. The maximum power density in wet H_2 was close to 0.5 W cm^{-2} and the power density in wet CH_4 is two times higher than in 5% H_2 , reaching a value of about 0.35 W cm^{-2} . The different slopes in the current–voltage plots under methane seem to suggest different reactions. After running a fuel cell for two days, cycling from 950 °C to 850 °C, in wet 5% H_2 , wet H_2 and wet methane, no traces of carbon could be detected visually or by thermal analysis. This anode exhibited a very fine microstructure, Fig. 4, with uniform particle size just less than 1 μm , an estimated 45% porosity and a clean interface with the electrolyte.

This work has demonstrated the concept of inducing functionality through disorder of extended defects. Through partially replacing titanium with some manganese and gallium, an SOFC anode with similar performance in hydrogen to nickel/zirconia cermets has been achieved. In contrast with these cermets, the electrode is remarkably active for the oxidation of CH_4 at high temperatures in the absence of excess steam; moreover, high OCVs reaching stable values between 1.2 V and 1.4 V have been obtained. Optimization of the microstructure allows a very marked improvement in the properties of the anode, but there remains a problem with fairly low lateral conductivity, which (especially for SOFC designs with long current paths) will require an additional current collector for practical application. Thus this material shows considerable promise as an electrochemical anode yielding the lowest polarization results so far reported for an oxide anode and excellent prospects for direct hydrocarbon use.

Received 16 September; accepted 1 November 2005.

- Kilner, J. A. & Steele, B. C. H. The effect of ion size on the energy of association between oxygen vacancies and dopant cations in oxide solid electrolytes. *J. Electrochem. Soc.* **129**, C143–C148 (1982).
- Irvine, J. T. S., Feighery, A. J., Fagg, D. P. & García-Martín, S. Structural studies on the optimisation of fast oxide ion transport. *Solid State Ionics* **136/137**, 879–885 (2000).
- Atkinson, A. et al. Advanced anodes for high-temperature fuel cells. *Nature Mater.* **3**, 17–27 (2004).
- Primdahl, S., Hansen, J. R., Grahl-Madsen, L. & Larsen, P. H. Sr-doped LaCrO_3 anode for solid oxide fuel cells. *J. Electrochem. Soc.* **148**, A74–A81 (2001).
- Singhal, S. C. & Kendall, K. *High Temperature Solid Oxide Fuel Cells: Fundamentals, Design and Applications* (Elsevier, Oxford, 2004).
- Matsuzaki, Y. & Yasuda, I. The poisoning effect of sulfur-containing impurity gas on a SOFC anode: Part I. Dependence on temperature, time and impurity concentration. *Solid State Ionics* **132**, 261–269 (2000).
- Zhu, W. Z. & Deevi, S. C. A review on the status of anode materials for solid oxide fuel cells. *Mater. Sci. Eng. A* **362**, 228–239 (2003).
- Park, S., Vohs, J. M. & Gorte, R. J. Direct oxidation of hydrocarbons in a solid-oxide fuel cell. *Nature* **404**, 265–267 (2000).

9. Perry, E., Tsai, T. & Barnett, S. A. A direct-methane fuel cell with a ceria-based anode. *Nature* **400**, 649–651 (1999).
10. Tao, S. W. & Irvine, J. T. S. A redox-stable efficient anode for solid-oxide fuel cells. *Nature Mater.* **2**, 320–323 (2003).
11. Canales-Vázquez, J., Tao, S. W. & Irvine, J. T. S. Electrical properties in $\text{La}_2\text{Sr}_4\text{Ti}_6\text{O}_{19.8}$: a potential anode for high temperature fuel cells. *Solid State Ionics* **159**, 159–165 (2003).
12. Marina, O. A., Canfield, N. L. & Stevenson, J. W. Thermal, electrical, and electrocatalytic properties of lanthanum-doped strontium titanate. *Solid State Ionics* **149**, 21–28 (2002).
13. Mukundan, R., Brosha, E. L. & Garzon, F. H. Sulfur tolerant anodes for SOFCs. *Electrochem. Solid-State Lett.* **7**, A4–A7 (2004).
14. Canales-Vázquez, J., Ruiz-Morales, J. C., Irvine, J. T. S. & Zhou, W. Sc-substituted oxygen excess titanates as fuel electrodes for SOFC. *J. Electrochem. Soc.* **152**, 1458–1465 (2005).
15. Marina, O. A. & Pederson, L. R. in *Proc. 5th European Solid Oxide Fuel Cell Forum* (ed. Huijsmans, J.) 481–489 (European Fuel Cell Forum, Oberrohrdorf, Switzerland, 2002).
16. Canales-Vázquez, J., Smith, M. J., Irvine, J. T. S. & Zhou, W. Studies on the reorganisation of extended defects with increasing n in the perovskite-based $\text{La}_{4}\text{Sr}_{n-4}\text{Ti}_n\text{O}_{3n+2}$ series. *Adv. Funct. Mater.* **15**, 1000–1008 (2005).
17. Holtappels, P., Bradley, J. L., Irvine, J. T. S., Kaiser, A. & Mogensén, M. Electrochemical characterization of ceramic SOFC anodes. *J. Electrochem. Soc.* **148**, A923–A929 (2001).
18. Irvine, J. T. S., et al. Optimisation of perovskite titanates and niobates as anode materials for SOFCs. in *Proc. 4th European SOFC Forum* (ed. McEvoy, A. J.) 471–477 (European Fuel Cell Forum, Oberrohrdorf, Switzerland, 2000).
19. Poeppelmeier, K. R., Leonowicz, M. E. & Longo, J. M. $\text{CaMnO}_{2.5}$ and $\text{Ca}_2\text{MnO}_{3.5}$ — new oxygen-defect perovskite-type oxides. *J. Solid State Chem.* **44**, 89–98 (1982).
20. McIntosh, S., Vohs, J. M. & Gorte, R. J. Effect of precious-metal dopants on SOFC anodes for direct utilization of hydrocarbons. *Electrochem. Solid-State Lett.* **6**, A240–A243 (2003).
21. Liu, J. & Barnett, S. A. Operation of anode-supported solid oxide fuel cells on methane and natural gas. *Solid State Ionics* **158**, 11–16 (2003).
22. Lin, Y., Zhan, Z., Liu, J. & Barnett, S. A. Direct operation of solid oxide fuel cell with methane fuel. *Solid State Ionics* **176**, 1827–1835 (2005).
23. Ovalle, A., Ruiz-Morales, J. C., Canales-Vázquez, J., Marrero-López, D. & Irvine, J. T. S. Mn-substituted titanates as efficient anodes for direct methane SOFCs. *Solid State Ionics* (submitted).

Acknowledgements This work was funded partly by a EU Marie Curie Fellowship and by EPSRC.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.T.S.I. (jtsi@st-and.ac.uk).

LETTERS

Spatially resolved observation of crystal-face-dependent catalysis by single turnover counting

Maarten B. J. Roelfsaers¹, Bert F. Sels¹, Hiroshi Uji-i², Frans C. De Schryver², Pierre A. Jacobs¹, Dirk E. De Vos¹ & Johan Hofkens²

Catalytic processes on surfaces have long been studied by probing model reactions on single-crystal metal surfaces under high vacuum conditions. Yet the vast majority of industrial heterogeneous catalysis occurs at ambient or elevated pressures using complex materials with crystal faces, edges and defects differing in their catalytic activity. Clearly, if new or improved catalysts are to be rationally designed, we require quantitative correlations between surface features and catalytic activity—ideally obtained under realistic reaction conditions^{1–3}. Transmission electron microscopy^{4–6} and scanning tunnelling microscopy^{7,8} have allowed *in situ* characterization of catalyst surfaces with atomic resolution, but are limited by the need for low-pressure conditions and conductive surfaces, respectively. Sum frequency generation spectroscopy can identify vibrations of adsorbed reactants and products in both gaseous and condensed phases⁹, but so far lacks sensitivity down to the single molecule level. Here we adapt real-time monitoring of the chemical transformation of individual organic molecules by fluorescence microscopy^{10–12} to monitor reactions catalysed by crystals of a layered double hydroxide immersed in reagent solution. By using a wide field microscope, we are able to map the spatial distribution of catalytic activity over the entire crystal by counting single turnover events. We find that ester hydrolysis proceeds on the lateral {10 $\bar{1}$ 0} crystal faces, while transesterification occurs on the entire outer crystal surface. Because the method operates at ambient temperature and pressure and in a condensed phase, it can be applied to the growing number of liquid-phase industrial organic transformations to localize catalytic activity on and in inorganic solids. An exciting opportunity is the use of probe molecules with different size and functionality, which should provide insight into shape-selective or structure-sensitive catalysis^{13–15} and thus help with the rational design of new or more productive heterogeneous catalysts.

The present study used [Li⁺-Al³⁺] layered double hydroxide (LDH) catalysts, which are composed of linked octahedrons of aluminium hydroxide. The octahedrons form gibbsite-type sheets stacked in a hexagonal space group, resulting in prismatic crystals with large basal {0001} planes¹⁶ (Fig. 1a). The occupation of two out of three octahedrons by Al³⁺ and the presence of Li⁺ in the residual octahedrons result in a positive charge on the layers. This feature endows the solid with anion exchange capacity, with exchanged anions such as OH⁻ located in the galleries between the sheets and at the gallery entrances located at the {10 $\bar{1}$ 0} crystal faces.

Because our experiments require large crystals with easily-identifiable faces, we used [Li⁺-Al³⁺] LDH, which readily forms such crystals. (The LDHs and their thermal treatment products used as adsorbents, drug delivery vehicles, polymer stabilizers and catalysts usually have other compositions^{17–20}.) Scanning electron microscopy on hydrothermally prepared LDH samples shows flat,

well-defined crystals with smooth basal planes. There are frequently intergrowths at the {0001} planes. Typically, the (0001)' plane of the intergrowing component is slightly tilted with respect to the (0001) plane of the underlying crystal component (Fig. 1b).

Processes such as molecular motor movements and enzyme dynamics have been visualized at the single-molecule level^{10,21} using fluorescent probes. Here we modify the recent work^{11,12} that monitored the hydrolytic activity of individual lipase B enzymes of *Candida antarctica* with the fluorogenic probe 5-carboxyfluorescein diacetate (C-FDA). Like other non-fluorescent esters of fluorescein such as fluorescein diacetate (FDA), C-FDA becomes emissive only upon catalytic hydrolysis in water-containing media, or upon catalytic transesterification with, for example, 1-butanol (Fig. 1a). In this study, a wide field fluorescence microscope was used to map the catalytic activity of the LDH crystals dispersed in milli-Q water (18 M Ω resistance) and deposited through spin coating on cleaned cover glasses. The cover glasses were then mounted at the bottom of a 1,000 μ l reaction chamber, which allows exposure of the LDH crystals to the fluorogenic reagent solution while monitoring the fluorescence signal of the product in inverted microscope mode (Fig. 1a).

In a first control experiment, an amine-functionalized glass slide was used to catalyse the reaction of C-FDA with 1-butanol. The glass surface was exposed for 1 h to a 1:1,000,000 mixture of N,N-dimethylaminopropyltrimethoxysilane (DMAPTS) and propyltrimethoxysilane (PTS) in CHCl₃, and the functionalized surface thoroughly rinsed (three times) with CHCl₃. For optimal observation of catalytic events, we used an excitation power of 4 kW cm⁻² as a good compromise between sensitivity and enhanced product bleaching. Upon addition of the non-fluorescent C-FDA precursor to the reaction chamber, bright spots appear owing to the formation of single molecules of emissive fluorescein (Fig. 2a). Photobleaching causes rapid disappearance of the spots (usually within 1 s) and thus prevents the rapid accumulation of background fluorescence. No spots were detected at all within the same time window when a clean glass slide replaced the catalyst, or when only propyl groups were anchored to the surface. Increasing the DMAPTS to PTS ratio during the glass surface preparation to 1:10,000 increases the concentration of catalytic sites, as evidenced by the larger density of fluorescent spots observed (Fig. 2b–d); quantification of the fluorescent signal suggests a 20-fold increase in the transesterification rate. When preparing this catalytic surface, part of the active surface monolayer was mechanically removed; in these defunctionalized zones, no spots are visible. Clearly, the formation of fluorescent molecules is directly related to the presence of catalytically active basic sites.

Next, LDH crystals were studied with the {0001} plane parallel to the cover glass. The catalytic activity of this class of catalysts is known to increase with surface area; based on combined CO₂ chemisorption

¹Microbial and Molecular Systems, Katholieke Universiteit Leuven, Kasteelpark Arenberg 23, B-3001 Leuven, Belgium. ²Division of Molecular and Nano Materials, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium.

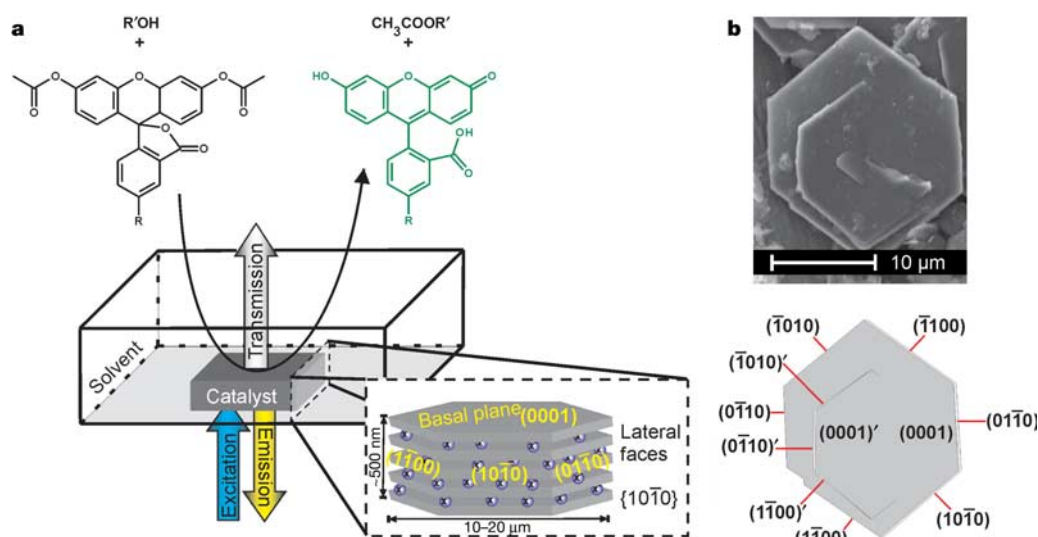


Figure 1 | Experimental set-up. **a**, Schematic drawing of the experimental set-up: the LDH particle is exposed to fluorescein ester ($R = -COOH$ for C-FDA; $R = -H$ for FDA) in $R'OH$ solution ($R' = -H$ for hydrolysis; $R' = -nC_4H_9$ for transesterification). A wide field microscope with 488 nm

excitation light was used. The inset shows the different crystallographic faces of a hexagonal LDH crystallite with indication of the Miller indices. **b**, Scanning electron micrograph of a typical LDH crystal with assignment of the different crystal faces for the intergrown crystal.

and kinetic catalytic data, it has been suggested that only a small fraction of the sites are responsible for the activity, and that these most active sites are probably situated at the edges and corners of the crystallites, or at the gallery entrances^{20,22,23}. Figure 1 illustrates catalytic sites located on the {0001} basal surface of the LDH crystal, and OH^- ions at the {1010} faces, which in the two-dimensional projection of the microscopic image appear at the edge of the hexagonal crystals. When carrying out transesterification of C-FDA in 1-butanol, spots appear all over the basal surface without preference for crystal edges (Fig. 3a, b, d). No fluorescence signal is detected in the solution surrounding the crystal, demonstrating that transesterification in this system is indeed catalysed by LDH in a true heterogeneous fashion. We also studied ester hydrolysis by using an aqueous reaction medium instead of an alcohol solvent (Fig. 1a), observing fluorescent spots primarily at the crystal edges and less at the basal surface (Fig. 3e). The fluorescence intensity distribution obtained when accumulating observations of the same crystal over 256 images clearly indicates that the hydrolysis activity follows the contours of the crystal (Fig. 3f, h). The activity of base catalysis by LDH is thus not always associated with the same type of sites: whereas transesterification occurs mainly at the {0001} plane, hydrolysis requires the {1010} faces where exchanged OH^- ions at the entrance of the galleries may be the active species. Such information has so far only been obtained very indirectly; for example, by correlating trends in catalytic activity with *ex situ* physical characterization of the catalyst samples used^{22,23}. Our *in situ* fluorescence approach, in contrast, has sufficient spatial resolution to allow us to simultaneously observe the activity associated with different crystal faces and thus uncover different reactive centres for two distinct base-catalysed reactions in the same crystal.

Time-dependent experiments permit determination of diffusion and reaction rates. For the reaction of 40 nM C-FDA with 1-butanol (Fig. 3a), counting the spots appearing on the basal plane reveals a typical transesterification rate of 4.2 molecules per $100 \mu m^2$ over a period of 96 ms, corresponding to $(7.2 \pm 0.4) \times 10^{-13} \text{ mol m}^{-2} \text{ s}^{-1}$ (Fig. 3c). With 600 nM C-FDA the number of spots increases and yields a rate of $(1.0 \pm 0.1) \times 10^{-11} \text{ mol m}^{-2} \text{ s}^{-1}$. Recording of rates for reagent concentrations between 0 and 700 nM (Fig. 3b) proves that in this range, the reaction is first order in reagent (data not shown). When FDA is used as the substrate for the transesterification reaction, a slightly lower rate of $(2.5 \pm 0.5) \times 10^{-13} \text{ mol m}^{-2} \text{ s}^{-1}$ for 100 nM FDA is measured (see Supplementary Video 3). We

attribute this difference to the extra carboxylate group of C-FDA, which leads to a higher concentration of adsorbed substrate at the positively charged surface, and hence to higher reaction rates. For the transesterification, the rates measured on {0001} are in fair agreement with transesterification rates measured for bulk LDH samples, which typically have surface areas between 50 and $250 \text{ m}^2 \text{ g}^{-1}$ (refs 24–27).

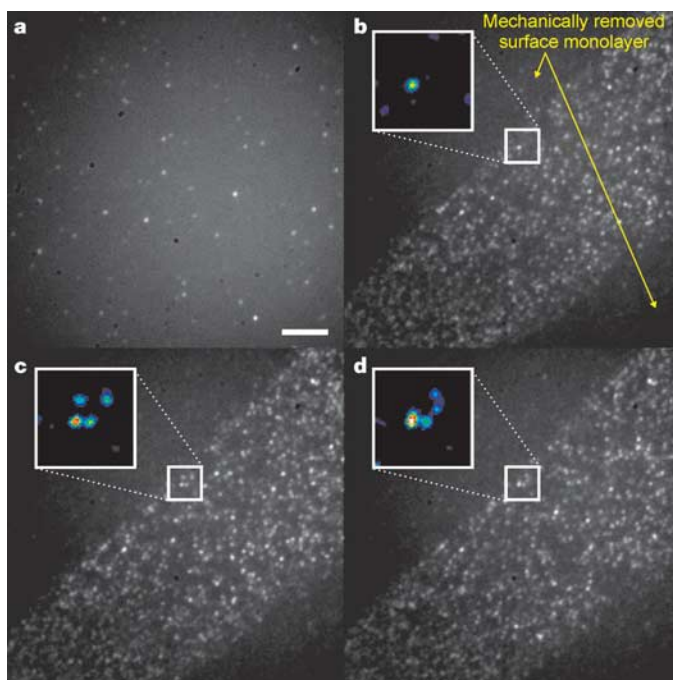


Figure 2 | *In situ* wide field fluorescence micrographs of C-FDA transesterification on propyl amine-functionalized cover glasses. **a**, Image of a cover glass functionalized with 1:1,000,000 DMAPTS:PTS mixture. Newly formed product molecules appear as bright spots, which are bleached soon after. Scale bar, 5 μm . **b**, Same reaction, on a cover glass prepared with 1:10,000 DMAPTS:PTS mixture. In the uniformly black zones, surface functionalization was mechanically removed. The insets show the formation of individual molecules in a $2.5 \times 2.5 \mu m^2$ square (see Supplementary Video 1). **c**, **d**, same as **b** but 96 and 192 ms later. Reactions were carried out at room temperature with 450 nM C-FDA in pure 1-butanol.

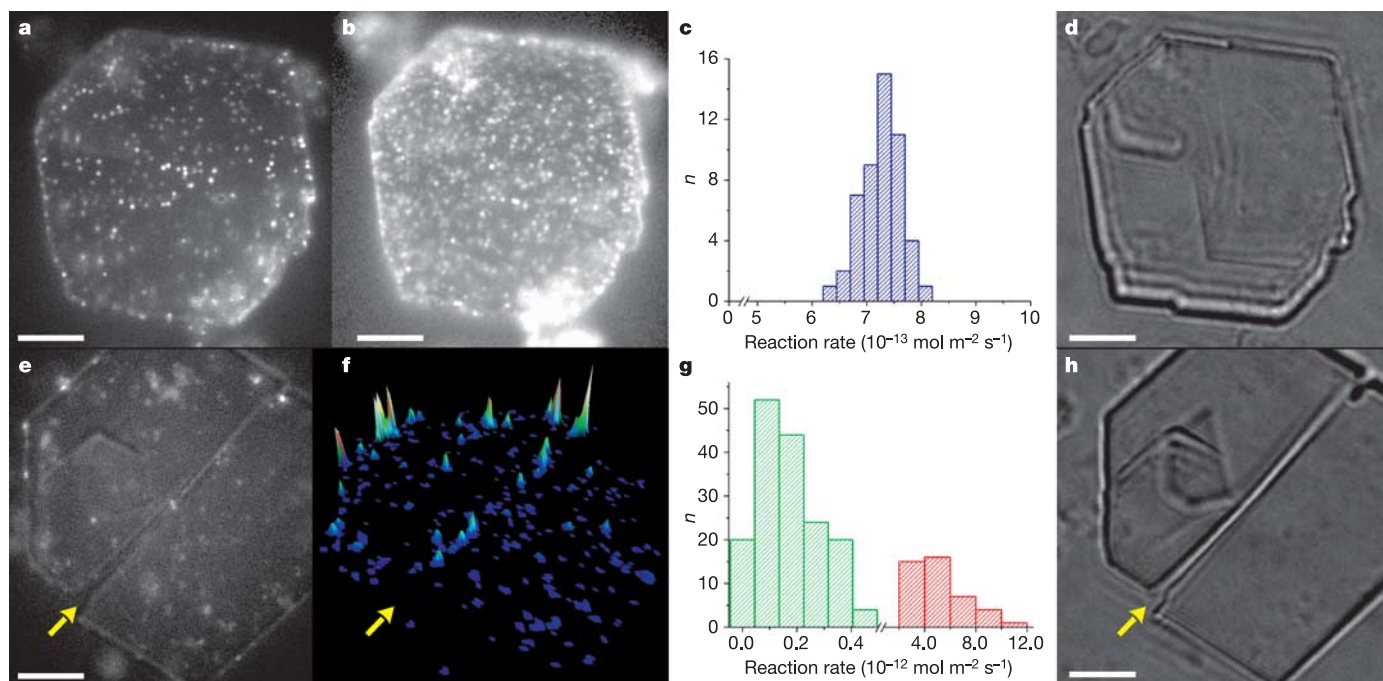


Figure 3 | Wide field images of catalytic reactions on individual LDH particles. **a–d**, Transesterification of C-FDA with 1-butanol at 40 nM (**a**) and 700 nM (**b**) ester concentration on the same LDH crystal (see Supplementary Video 2). **c**, Distribution of initial reaction rates for $1 \mu\text{m}^2$ domains on the crystal faces ($n = 50$). **d**, Transmission image of the crystal. **e–h**, Hydrolysis of 600 nM C-FDA on an LDH crystal. **e**, Fluorescence image, showing formation of single product molecules mainly at crystal edges (96 ms per image). **f**, Accumulated spot intensity on the same crystal over

256 consecutive images. **g**, Distribution of initial reaction rates for $1 \mu\text{m}^2$ domains on the faces of the LDH crystal ($n = 207$). The distribution clearly shows two statistically different subpopulations. The fast population corresponds to active domains located on the $\{10\bar{1}0\}$ faces (red), whereas the $\{0001\}$ faces host the slow population (green). **h**, Transmission image. Scale bars, 5 μm . Yellow arrows in **e**, **f**, **h** indicate the same viewing direction on the same crystal.

For the hydrolysis of C-FDA (600 nM) on an LDH catalyst particle, the contributions of the different crystallographic planes can be clearly separated (Fig. 3g). Spots formed on the $\{10\bar{1}0\}$ faces of either the main crystal or intergrown crystals correspond to at least 85% of the overall activity of the particle. On the $\{0001\}$ plane, the hydrolysis rate is only $(1.7 \pm 1.2) \times 10^{-13} \text{ mol m}^{-2} \text{ s}^{-1}$ for 600 nM C-FDA, while at the $\{10\bar{1}0\}$ faces the rate averages to $(4.7 \pm 2.3) \times 10^{-12} \text{ mol m}^{-2} \text{ s}^{-1}$. Counting of single turnovers thus allows us to quantify the heterogeneity in reaction kinetics at the level of individual crystal faces, thus providing important information that could aid the design of industrial catalysts with improved activity.

Close inspection of the Supplementary videos reveals motion of the fluorescent dots. This allows us to monitor the diffusion of the fluorescent products of the transesterification reaction along the large $\{0001\}$ crystal face over several time frames, before bleaching occurs (Fig. 4). The displacements $|r|$ travelled between the observations by many product molecules can be adequately fitted as $f(r) = (2r/\langle r^2 \rangle) \exp[-r^2/\langle r^2 \rangle]$, with $\langle r^2 \rangle$ the mean square displacement²⁸. From this a diffusion coefficient, D , can be calculated, based on $D = \langle r^2 \rangle / 4t$ and with $t = 96 \text{ ms}$. We find that over 90% of the 5-carboxyfluorescein product molecules are highly mobile, with $D = (3.0 \pm 0.5) \times 10^{-14} \text{ m}^2 \text{ s}^{-1}$. A small fraction of the product molecules is almost immobile, with $D \leq 1 \times 10^{-16} \text{ m}^2 \text{ s}^{-1}$. The diffusion of dye molecules trapped in mesoporous siliceous materials exhibits a similar bimodal behaviour, with fast-moving as well as stationary molecules²⁸. In the case of fluorescein dye, a diffusion coefficient of $D = (3.0 \pm 0.5) \times 10^{-14} \text{ m}^2 \text{ s}^{-1}$ was measured for more than 80% of the molecules, while $D \leq 7 \times 10^{-16} \text{ m}^2 \text{ s}^{-1}$ for the rest of the molecules. Catalyst deactivation by strong product adsorption is known in basic catalysis, and we are currently investigating whether the immobile molecules are related to this phenomenon.

Extending and improving the use of *in situ* fluorescence microscopy as a tool for probing heterogeneous catalysis calls for better spatial resolution, and for the availability of probe molecules that can be used to monitor a wide range of different chemical reactions. We anticipate that spatial resolution might be improved by an order of magnitude by future advances in fluorescence microscopy, such as STED (stimulated emission-depletion) methods²⁹. We have also already identified suitable probes for base-, acid- and metal-catalysed reactions, such as ester exchange, Friedel-Crafts or hydrogenation reactions. In a first extension of the present work, we have mapped the spatial distribution of Friedel-Crafts catalytic activity in the channels of a mordenite zeolite crystal (see Supplementary Information). We anticipate that reporter probes with different size and functionality could be used to gain insight into steric, electronic and polarity characteristics of reactive sites. We expect that such developments, used in conjunction with methods for characterizing the inorganic catalyst material, will provide better insight into challenging aspects of heterogeneous catalysis, such as shape-selective¹³ and key-lock³⁰ catalysis in zeolites, or structure-sensitive catalysis on supported metals^{14,15}.

METHODS

Materials. $[\text{LiAl}_2(\text{OH})_6]^+\text{OH}^- \cdot n\text{H}_2\text{O}$ was hydrothermally synthesized following a reported procedure³¹. A toluene solution of 35 μmol Al tri-*sec*-butoxide was added dropwise to an aqueous solution of LiOH (Al/Li = 1.5). Gels were treated hydrothermally for 10 d at 403 K. Products were washed with hot water to eliminate the excess lithium salts. Crystallinity and crystal size were checked using scanning electron microscopy and X-ray diffraction (hexagonal unit cell parameters: $a = 5.09 \text{ \AA}$ and $c = 15.2 \text{ \AA}$). Pure, large hexagonal LDH crystals were obtained, with smooth basal planes about 10–20 μm in diameter and a characteristic thickness of about 500 nm. As only OH^- ions are available during synthesis, the LDH is completely in the OH^- form.

Observation of catalytic processes. The wide field microscope consists of an inverted microscope (IX-71, Olympus) with a 100 $\times$, 1.3 NA oil immersion

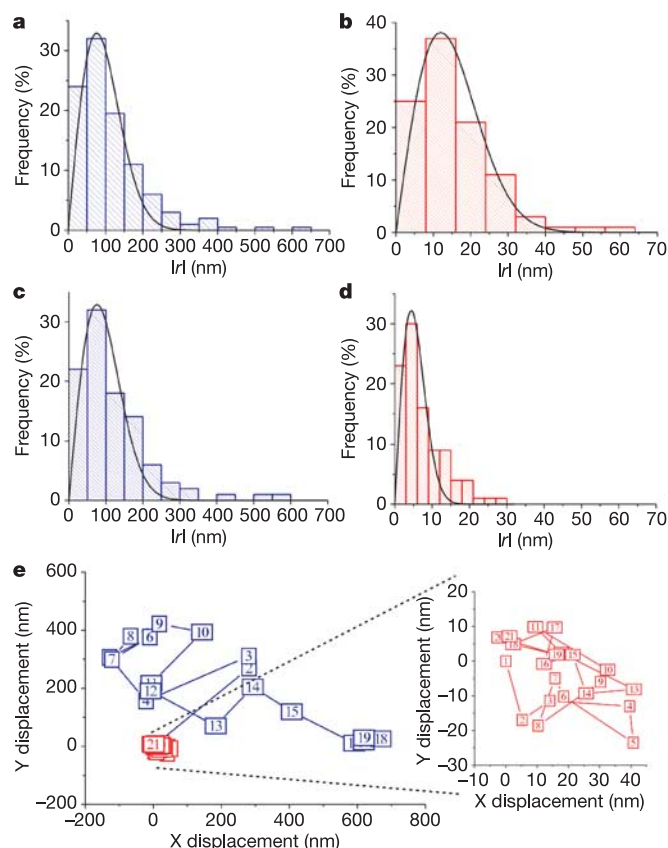


Figure 4 | Analysis of single product molecules randomly diffusing on (0001) LDH surfaces after transesterification. The time lag between the frames is 96 ms (see Supplementary Video 4). **a, b,** The distribution ($n = 100$) of step sizes r for fluorescein shows two different populations. **a,** Over 80% of the molecules show a high mobility with a mean displacement of more than 100 nm. **b,** For a small group of molecules, the displacement is of the same order as the positioning accuracy (~ 10 nm). **c, d,** As in **a, b,** but with 5-carboxyfluorescein, produced from C-FDA. **e,** The trajectories of two representative fluorescein molecules (red: stationary; blue: highly mobile; same colours as histograms in **a** and **b**). Numbers in **e** represent consecutive positions of the molecules. Reactions were carried out at room temperature with 40 nM C-FDA or 100 nM FDA in pure 1-butanol.

objective lens and a highly sensitive cooled Electron Multiplying-CCD (cascade 512B, Princeton Instruments Inc.). The wide field illumination was achieved by focusing the expanded circular polarized 488 nm light from an Ar⁺ laser (Stabilite 2017, Spectra-Physics) onto the back focal plane of the objective. Emission is collected by the same objective and imaged by the CCD after passing through a dichroic mirror and spectral filters removing the excitation light. The image was expanded 3.3 × before the CCD, resulting in a field of view of 24.2 × 24.2 μm. Transmission images were obtained by Köhler illumination.

Received 14 June; accepted 8 November 2005.

- Thomas, J. M. & Thomas, W. J. *Principles and Practice of Heterogeneous Catalysis* (Wiley-VCH, Weinheim, 1996).
- Thomas, J. M. Catalysis and surface science at high resolution. *Faraday Discuss.* **105**, 1–31 (1996).
- Ertl, G., Knözinger, H. & Weitkamp, J. *Handbook of Heterogeneous Catalysis* (Wiley-VCH, Weinheim, 2000).
- Hansen, T. W. *et al.* Atomic-resolution in situ transmission electron microscopy of a promoter of a heterogeneous catalyst. *Science* **294**, 1508–1510 (2001).
- Hansen, P. L. *et al.* Atom-resolved imaging of dynamic shape changes in supported copper nanocrystals. *Science* **295**, 2053–2055 (2002).

- Helveg, S. *et al.* Atomic-scale imaging of carbon nanofibre growth. *Nature* **427**, 426–429 (2004).
- Zambelli, T., Barth, J. V., Wintterlin, J. & Ertl, G. Complex pathways in dissociative adsorption of oxygen on platinum. *Nature* **390**, 495–497 (1997).
- Wolff, J., Papathanasiou, A. G., Kevrekidis, I. G., Rotermund, H. H. & Ertl, G. Spatiotemporal addressing of surface activity. *Science* **294**, 134–137 (2001).
- Dellwig, T., Rupprechter, G., Unterhalt, H. & Freund, H. J. Bridging the pressure and materials gap: High pressure sum frequency generation study on supported Pd nanoparticles. *Phys. Rev. Lett.* **85**, 776–779 (2000).
- Lu, H. P., Xu, L. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282**, 1877–1882 (1998).
- Velonia, K. *et al.* Single-enzyme kinetics of CALB-catalyzed hydrolysis. *Angew. Chem. Int. Edn Engl.* **44**, 560–564 (2005).
- Flomenbom, O. *et al.* Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. *Proc. Natl Acad. Sci. USA* **102**, 2368–2372 (2005).
- Corma, A. Inorganic solid acids and their use in acid-catalyzed hydrocarbon reactions. *Chem. Rev.* **95**, 559–614 (1995).
- Boudart, M., Aldag, A., Benson, J. E., Dougharty, N. A. & Harkins, C. G. On the specific activity of platinum catalysts. *J. Catal.* **6**, 92–99 (1966).
- Bernasek, S. L., Siekhaus, W. J. & Somorjai, G. A. Molecular-beam study of hydrogen-deuterium exchange on low- and high-Miller-index platinum single-crystal surfaces. *Phys. Rev. Lett.* **30**, 1202–1204 (1973).
- Fogg, A. M., Freij, A. J. & Parkinson, G. M. Synthesis and anion exchange chemistry of rhombohedral Li/Al layered double hydroxides. *Chem. Mater.* **14**, 232–234 (2002).
- Cavani, F., Trifiro, F. & Vaccari, A. Hydrotalcite-type anionic clays: preparation, properties and applications. *Catal. Today* **11**, 173–301 (1991).
- Cai, H., Hillier, A. C., Franklin, K. R., Nunn, C. C. & Ward, M. D. Nanoscale imaging of molecular adsorption. *Science* **266**, 1551–1555 (1994).
- Sels, B. *et al.* Layered double hydroxides exchanged with tungstate as biomimetic catalysts for mild oxidative bromination. *Nature* **400**, 855–857 (1999).
- Sels, B. F., De Vos, D. E. & Jacobs, P. A. Hydrotalcite-like anionic clays in catalytic organic reactions. *Catal. Rev.* **43**, 443–488 (2001).
- Noji, H., Yasuda, R., Yoshida, M. & Kinosita, K. Direct observation of the rotation of F1-ATPase. *Nature* **386**, 299–302 (1997).
- Roelofs, J. C. A. A., Lensveld, D. J., van Dillen, A. J. & de Jong, K. P. On the structure of activated hydrotalcites as solid base catalysts for liquid-phase aldol condensation. *J. Catal.* **203**, 184–191 (2001).
- Abelló, S. *et al.* Aldol condensations over reconstructed Mg–Al hydrotalcites: structure–activity relationships related to the rehydration method. *Chem. Eur. J.* **11**, 728–739 (2005).
- Corma, A., Iborra, S., Miquel, S. & Prim, J. Production of food emulsifiers, monoglycerides, by glycerolysis of fats with solid base catalysts. *J. Catal.* **173**, 315–321 (1998).
- Engel, D. J., Malloy, T. P., & Nickl, P. K. Transesterification using metal oxide solid solutions as the basic catalyst. US Patent 5,350,879 (1993).
- Watanabe, Y. & Tatsumi, T. Hydrotalcite-type materials as catalysts for the synthesis of dimethyl carbonate from ethylene carbonate and methanol. *Micropor. Mesopor. Mater.* **22**, 399–407 (1998).
- Abelló, S. *et al.* Aldol condensations over reconstructed Mg–Al hydrotalcites: structure–activity relationships related to the rehydration method. *Chem. Eur. J.* **11**, 728–739 (2005).
- Hellriegel, C., Kirstein, J. & Bräuchle, C. Tracking of single molecules as a powerful method to characterize diffusivity of organic species in mesoporous materials. *N. J. Phys.* **7**, 1–14 (2005).
- Hell, S. W. Toward fluorescence nanoscopy. *Nature Biotechnol.* **21**, 1347–1355 (2003).
- Martens, J. A. *et al.* Evidences for pore mouth and key–lock catalysis in hydroisomerization of long n-alkanes over 10-ring tubular pore bifunctional zeolites. *Catal. Today* **65**, 111–116 (2001).
- Serna, C. J., Rendon, J. L. & Iglesias, J. E. Crystal-chemical study of layered $[Al_2Li(OH)_6]^+X^- \cdot nH_2O$. *Clays Clay Miner.* **30**, 180–184 (1982).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.B.J.R. thanks the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for a fellowship, B.F.S. thanks the FWO-Flanders for a post-doctoral fellowship and the KUL for a guest professor position. This work was performed within the framework of the IAP-V-03 programme ‘Supramolecular Chemistry and Catalysis’ of the Belgian Federal government and of GOA-2/01.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.E.D.V. (dirk.devos@biw.kuleuven.be) or J.H. (johan.hofkens@chem.kuleuven.be).

LETTERS

Malaria early warnings based on seasonal climate forecasts from multi-model ensembles

M. C. Thomson¹, F. J. Doblas-Reyes², S. J. Mason¹, R. Hagedorn², S. J. Connor¹, T. Phindela³, A. P. Morse⁴ & T. N. Palmer²

The control of epidemic malaria is a priority for the international health community and specific targets for the early detection and effective control of epidemics have been agreed¹. Interannual climate variability is an important determinant of epidemics in parts of Africa² where climate drives both mosquito vector dynamics and parasite development rates³. Hence, skilful seasonal climate forecasts may provide early warning of changes of risk in epidemic-prone regions. Here we discuss the development of a system to forecast probabilities of anomalously high and low malaria incidence with dynamically based, seasonal-timescale, multi-model ensemble predictions of climate, using leading global coupled ocean–atmosphere climate models developed in Europe. This forecast system is successfully applied to the prediction of malaria risk in Botswana, where links between malaria and climate variability are well established⁴, adding up to four months lead time over malaria warnings issued with observed precipitation and having a comparably high level of probabilistic prediction skill. In years in which the forecast probability distribution is different from that of climatology, malaria decision-makers can use this information for improved resource allocation.

Malaria is the most important parasitic infection in people, accounting for an estimated 500 million clinical attacks worldwide and more than 1 million deaths a year, mostly in sub-Saharan Africa⁵. On this continent, epidemic malaria is associated with seasonally warm semi-arid and highland areas where approximately 124 million people are considered at risk of climate-related malaria epidemics⁶. For improved control in epidemic regions, the World Health Organisation (WHO) advocates the development of integrated malaria early warning systems based on vulnerability assessment, seasonal climate forecasts, weather and environmental monitoring and case surveillance^{2,7}.

In the semi-arid country of Botswana, malaria is ranked as one of the major public health problems and consequently there is a well developed National Malaria Control Programme that supports a set of key control activities: (1) vector control through annual indoor residual house spraying in endemic areas and in response to epidemics, (2) chemoprophylaxis for target groups during the malaria season, and (3) prompt and effective case management including drug therapy to affected individuals. After the devastating regional malaria epidemic of 1996, malaria control was reinvigorated through a range of initiatives including the development of an early warning system for epidemics, which peak in March and April following the bulk of the rainy season (November–February, NDJF)⁴. As a notifiable disease, malaria cases are reported routinely to the Ministry of Health and epidemic alerts are based on whether weekly numbers of cases exceed pre-defined thresholds. An earlier study⁴ has indicated that December–February seasonally averaged monitored rainfall and

also sea surface temperature (SST) for the same time period provide significant predictive skill for the malaria season one month in advance of its seasonal peak. Using data on yearly malaria incidence in Botswana, we discuss whether the lead time for providing reliable early warnings using observed precipitation data (for improved targeting of insecticides, drug stocks and other control measures) can be increased significantly using skilful seasonal climate forecasts.

The results in this Letter are based on re-forecasts (retrospective forecasts) using models that form the basis of a future real-time operational multi-model ensemble forecast system developed out of the original DEMETER (Development of a European Multi-model Ensemble Forecast System for Seasonal to Interannual Climate Prediction) project⁸. DEMETER used a multi-model ensemble forecast system for seasonal-to-interannual climate variability, comprising state-of-the-art coupled ocean–atmosphere global climate models developed in Europe.

An ensemble climate forecast system predicts not only the most likely evolution of climate, but also the uncertainty in such a prediction. More generally, an ensemble climate forecast system

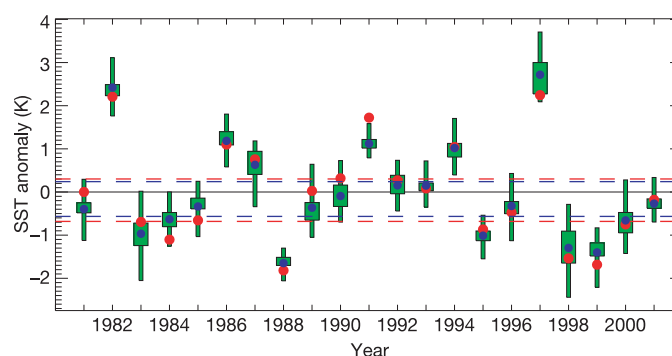


Figure 1 | Time series of the one-month lead December–February predicted SST averaged over the tropical Pacific region Niño3.4. The boundaries of Niño3.4 are 5° N–5° S and 170° W–120° W. The multi-model ensemble spread is depicted by the green box-and-whisker representation, the whiskers containing one-third of the ensemble members. The blue dots represent the ensemble mean and the larger red bullets show the ERA-40 anomalies. The horizontal dotted lines indicate the ERA-40 (red) and multi-model re-forecast (blue) climatological tercile boundaries. The ensemble-mean correlation with the reference is 0.97, while the ROC score (see text) for the three tercile categories (above normal, normal and below normal) ranges between 0.97 and 1.00. By comparison, the ROC scores for SST predictions based on the persistence of the August–October ERA-40 anomalies take values of the order of 0.85.

¹International Research Institute for Climate and Society (IRI), The Earth Institute at Columbia University, New York, New York 10964-8000, USA. ²European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, RG2 9AX, Reading, UK. ³National Malaria Control Programme, Epidemiology and Disease Control Unit, Ministry of Health, Private Bag 00269, Gaborone, Botswana. ⁴Department of Geography, University of Liverpool, Liverpool, L69 7ZT, UK.

Table 1 | ROC areas for annual standardized log malaria incidence predictions

Event	CMAP	DEMETER
Very high	0.93 (0.70–1.00)	0.80 (0.59–0.96)
High	0.36 (0.06–0.70)	0.48 (0.18–0.79)
Low	0.33 (0.11–0.58)	0.34 (0.12–0.60)
Very low	0.91 (0.73–1.00)	0.83 (0.61–0.99)

The predictions are carried out using observed CMAP and DEMETER multi-model ensemble precipitation over NDJF and have similar forecast quality. The four events for values are described in the Methods section. CMAP predictions are assessed as probabilistic forecasts with fixed, but unknown uncertainty²⁴. The 95% confidence limits obtained by bootstrapping with a sample size of 10,000 are shown in parentheses. CMAP-based predictions are available at the beginning of March, but the DEMETER multi-model predictions would be available to users four months earlier, at the beginning of November, a potentially crucial increase in lead time for the management of malaria epidemics.

predicts a probability distribution of climate. The uncertainties in climate prediction that generate such probability distributions arise from imprecise initial conditions, and numerical approximations to the underlying partial differential equations that govern climate⁹. In the DEMETER ensemble forecast system, model uncertainty is represented by the fact that the sub-grid parameterizations in the component models have, to a large extent, been developed independently. Here we analyse data from six-month integrations started on 1 November of every year (see Methods).

The physical basis for seasonal climate prediction lies, to great

extent, in the predictability of slowly varying interactions between the atmosphere and the oceans. The prototypical phenomenon illustrating such predictability is El Niño/Southern Oscillation (ENSO) which has been shown to be predictable on seasonal time-scales^{10,11}. It is known that interannual variability in the strength of ENSO can strongly influence the climate and maize yields of southern Africa¹² and is also a predictor for unusually high or low malaria anomaly years in Botswana⁴. To illustrate the ability of the DEMETER system to forecast atmosphere–ocean interactions, Fig. 1 shows probabilistic predictions of the SST during December–February averaged over the region in the tropical Pacific known as Niño3.4. These probabilistic forecasts are the most skilful dynamically based predictions reported to date, and substantially outperform deterministic predictions based on persistence of the SST anomalies of the month previous to the dynamical forecast start date¹³.

We now focus on the DEMETER prediction of the climate of southern Africa, a necessary condition for the accurate prediction of malaria incidence. Cumulative precipitation anomalies for NDJF using CMAP-estimated precipitation data (see Methods) are shown in Fig. 2a and b composited for the five highest and lowest malaria incidence anomaly years of the data sample (1982–2002). These figures illustrate that higher than expected malaria years are associated with above-average precipitation, while the lowest malaria years are associated with below-average precipitation. In addition, Fig. 2c

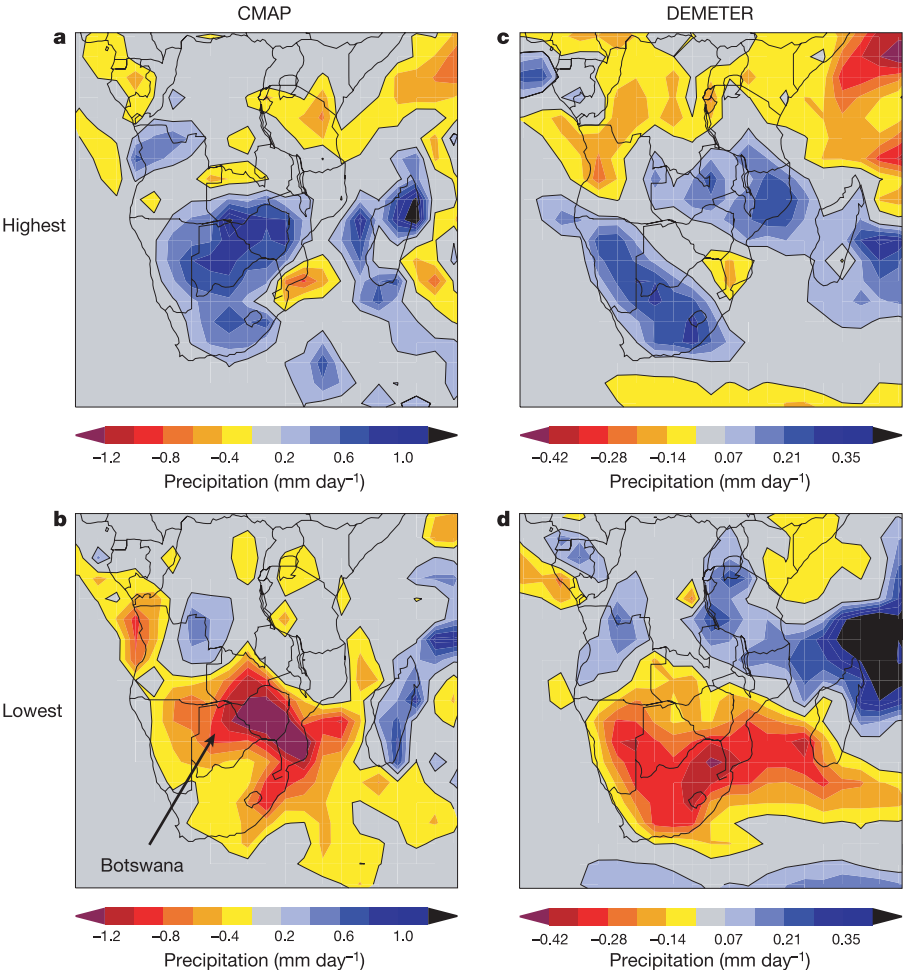


Figure 2 | Composites of austral summer precipitation for central and southern Africa as a function of the standardized malaria annual incidence for Botswana. Mean precipitation anomalies for the five years with the highest (a, c) and lowest (b, d) malaria annual incidence in Botswana for November–February CMAP (a, b) and DEMETER (c, d) ensemble-mean

precipitation. A different colour scale was used for the observed (CMAP) and re-forecast (DEMETER) precipitation because of the unavoidable reduction in amplitude of the predicted anomalies when averaging the 27 ensemble members.

and d shows that during the high and low malaria years, the one-to-four-month NDJF ensemble-mean precipitation from the DEMETER forecasts (available at the beginning of November) are above and below average over much of southern Africa, respectively.

The association between malaria incidence and precipitation depicted in Fig. 2a and b can be made more precise if we use a quadratic relationship similar to that described in an earlier study⁴ between seasonal precipitation averaged over Botswana and the logarithm of standardized (detrended) confirmed malaria incidence (per 1,000 of population), as illustrated in Fig. 3. This nonlinear relationship provides the quantitative link that allows the DEMETER probability forecasts of precipitation to be transformed into probability forecasts of malaria incidence. The quadratic relationship can also be used to issue malaria warnings based on monitored precipitation⁴, such as NDJF CMAP precipitation, although this would be available only at the beginning of March, four months later than the multi-model forecast issue date. As an illustration, Fig. 4 shows DEMETER forecast probability distributions of malaria incidence associated with a year of very high malaria incidence, and another year of very low malaria incidence. For decision-makers, that these probability distributions are not sharp quantifies the inevitable uncertainties that exist in forecasting a variable that is strongly influenced by partially chaotic climatic processes. On the other hand, forecast probability distributions for the high malaria years are distinctly different from the probability distributions for the low malaria years: the null-hypothesis that the distributions for DEMETER-based predictions in the two categories are not distinct can be rejected using a Kolmogorov–Smirnov test with a *P* value smaller than 0.0001.

The quality of these forecast probability distributions of malaria incidence has been assessed using the area under the Receiver Operating Characteristic (ROC) curve¹⁴ for the events ‘values higher than the top quartile’ (very high), ‘between the top quartile and the median’ (high), ‘between the median and the lowest quartile’ (low) and ‘lower than the lowest quartile’ (very low). The ROC score is an appropriate measure for assessing the value of probability climate forecasts for decision-making^{14,15}. Suppose it is decided to take precautionary action (targeting drug or insecticide supplies to a particular region) if the probability of malaria in the highest-quartile category exceeds some chosen threshold probability. Over the full forecast data sample, the number of correct and incorrect decisions

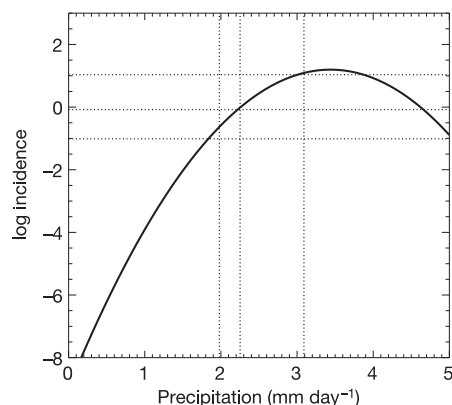


Figure 3 | Relationship between standardized log malaria annual incidence and summer precipitation for Botswana. Standardized log malaria annual incidence (per 1,000 population) versus November–February CMAP precipitation for the period 1982–2002 as obtained from fitting a quadratic function to observations⁴. The quadratic function is $y = -8.9 + 5.9x - 0.8x^2$, where x denotes precipitation (measured in mm day^{-1}) and y the standardized log malaria incidence index. The horizontal and vertical dotted lines denote the quartiles of the standardized log malaria incidence index, and precipitation, respectively.

can be assessed for this particular threshold probability: these are referred to in terms of ‘hit rate’ and ‘false-alarm rate’, respectively. The ROC score is a measure of the mean hit rate to false-alarm rate, averaging over all threshold probabilities. A ROC score larger than 0.5 indicates a forecast system more skilful than a forecast based on climatological probabilities; a ROC score of 1.0 indicates a perfect forecast system.

Table 1 shows the ROC scores for predictions of standardized malaria incidence, first using the CMAP precipitation and second with the DEMETER multi-model ensemble forecasts initialized on 1 November. For forecasts of both very low and very high malaria incidence, the multi-model DEMETER forecasts have significantly positive ROC scores. The DEMETER scores are slightly lower than the CMAP scores, but we point out that the DEMETER forecasts are available at least four months before the CMAP data becomes available in early March.

Our evidence supports the use of multi-model ensemble climate predictions initialized at least five months before the peak malaria season in Botswana and four months earlier than a prediction based on monitored precipitation for early warning of epidemic malaria risks. The significant gain in lead time with small reduction in prediction skill suggests that probabilistic seasonal climate forecasts can be used to predict malaria incidence, not only in Botswana, but also in neighbouring epidemic-prone areas, as shown in Fig. 2. These results are not only of importance to the malaria control managers in the region who are actively involved in using climate information to achieve malaria-reduction targets¹⁶, but are also relevant to other resource managers (health, hydrology, agriculture, and so on) faced with climate-sensitive decisions. Unlike the products that would be derived from single forecast models, the malaria incidence forecasts described here allow the user to obtain a probability distribution of disease risk, thus indicating to the user the uncertainty of the information available in the forecast and the extent to which it differs from climatology.

Although the greatest burden of malaria in Africa is suffered by those living in endemic regions, epidemics pose a serious threat to many millions of people⁶ and their prevention remains a priority¹. The ability of a malaria early-warning system to improve resource allocation and assist in the reduction of malaria morbidity and mortality depend on decision-makers’ capacity to make effective use of new information within their own control paradigm. To this end we have focused our efforts on the integration of the multi-model DEMETER forecast system into the routine epidemic malaria control activities currently promoted by the WHO-AFRO Southern Africa Inter-Country Malaria Team (SAMC) in Zimbabwe and undertaken by the National Malaria Control Unit in Botswana. It

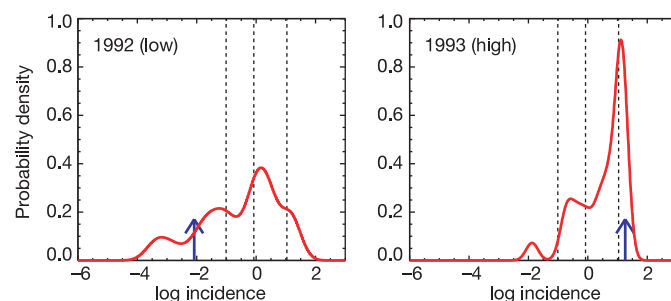


Figure 4 | Forecast probability distribution function of standardized log malaria annual incidence for Botswana. The probability distribution functions of predicted standardized log malaria annual incidence for the years 1992 (anomalously low incidence, left) and 1993 (anomalously high incidence, right) computed with the DEMETER multi-model ensemble forecast system are depicted in red. The vertical dashed lines denote the quartiles of the distribution of the standardized log malaria incidence index and the vertical blue arrows indicate the values recorded by the Botswana Ministry of Health.

is here that the final arbiters of the utility of the forecasts will be found.

METHODS

Malaria incidence data. Botswana has compiled national annual data on cases of laboratory-confirmed malaria incidence for the period 1982–2002. This data set offers a unique opportunity to analyse and predict malaria incidence in a desert-fringe area. Confirmed incidence per 1,000 population was modelled⁴ to remove non-climate trends, largely attributed to chloroquine resistance¹⁷ and the impact of a major policy intervention, which included the change of the first-line drug. The residual, a standardized log malaria incidence index, is the variable used in this study. Epidemiological and population data were obtained from the Ministry of Health's Epidemiology and Disease Control Unit, and the Central Statistics Office in the capital city, Gaborone.

Precipitation data. Precipitation in Botswana is concentrated in the period November to March and is subject to high interannual variability¹⁸. For this analysis, the 2.5°-resolution monthly data from the Climate Prediction Center Merged Analysis of Precipitation (CMAP)¹⁹ were averaged across the 20 grid points between 17.5°–27.5° S and 17.5°–30.0° E for the rainy season before the peak malaria season (NDJF) from 1979 to date. Calendar years for precipitation are referred to by the year of the January used in the seasonal average. The best relationship between seasonally averaged precipitation is described by the quadratic model depicted in Fig. 3 (ref. 4).

DEMETER climate predictions. The DEMETER multi-model ensemble system for climate prediction comprises seven global coupled ocean–atmosphere models^{8,20}. Re-forecasts have been produced over the period 1958–2001, although the period common to the seven models is 1980–2001. The three models used in this study comprise the operational DEMETER system and are from the following institutions: the ECMWF (European Centre for Medium-Range Weather Forecasts), Centre National de Recherches Météorologiques (Météo-France, France), and The Met Office (UK). The corresponding re-forecasts have a common period of 1959–2001. The uncertainty in the initial conditions derives from the fact that each DEMETER model is integrated from nine different initial conditions, each with different estimates of initial-condition uncertainty. Atmospheric and land-surface initial conditions are taken directly from the ECMWF re-analysis ERA-40 (ref. 21). The nine-member ensemble re-forecasts from each model were integrated for 180 days. Because of the computational cost of running multi-model ensembles, DEMETER re-forecast start dates were quarterly rather than monthly: 1 February, 1 May, 1 August and 1 November. More detailed information on the models and the initialization procedures can be found on <http://www.ecmwf.int/research/demeter/general/docmodel/index.html>. According to the set of available start dates, predictions for NDJF (months one to four of the simulations) were used. DEMETER multi-model ensemble predictions were averaged over Botswana in the same manner as the CMAP data.

Climate prediction calibration. Given the typical systematic errors of coupled model simulations, each prediction is corrected to have statistical properties similar to those of the observed precipitation. To allow robust estimates of the correction, the CMAP precipitation time series for Botswana starting in 1979 has been extended back to 1959 using the Climate Research Unit of the University of East Anglia (CRU) precipitation data^{22,23}, which is highly correlated with CMAP. This also limits the number of coupled models available from DEMETER to the three mentioned above. A scheme that mimics an operational prediction system has been used to correct the precipitation ensemble predictions. Mean and variance estimates for a specific year are obtained for each single model and for CMAP/CRU for the period from 1960 to the year before the target year. The difference in means and ratio of variances between each model and the reference is then computed and applied to the predictions of the target year. The same procedure is used for subsequent years. This scheme makes available each year a longer sample to compare the CMAP/CRU data with each single model. The corrected DEMETER-predicted precipitation was used to create ensemble predictions of standardized log malaria incidence for the period 1982–2002. The quadratic relationship between CMAP precipitation and malaria incidence is applied to the corrected predicted precipitation of each ensemble member to obtain the malaria incidence predictions.

Formulation of probability forecasts. Probabilistic predictions have been formulated for four different categories defined using quartiles: very high, high, low, and very low (see text). The most useful for assessing malaria epidemic risk are the very low and high categories, defined by the values below and above the first and third quartile, respectively. Probabilities have been estimated as the relative number of members of a given ensemble of predicted standardized log malaria incidence falling within a given category, a simple method that assigns the same weight to every single model.

Received 6 May; accepted 1 December 2005.

1. WHO/UNICEF. *The World Malaria Report*. 1–120, Report no. WHO/HTM/MAL/2005.1102 (WHO/UNICEF, Geneva, 2005); (<http://rbm.who.int/wmr2005/index.html>).
2. Najera, J. A., Kouznetsov, R. L. & Delacollette, C. *Malaria Epidemics: Detection And Control, Forecasting And Prevention*. 1–81, Report no. WHO/MAL/98.1084 (WHO, Geneva, 1998); (http://www.who.int/malaria/docs/najera_epidemics/naj_toc.htm).
3. Thomson, M. C., Connor, S. J., Milligan, P. J. M. & Flasse, S. P. The ecology of malaria—as seen from Earth-observation satellites. *Ann. Trop. Med. Parasitol.* **90**, 243–264 (1996).
4. Thomson, M. C., Connor, S. J., Phindela, T. & Mason, S. J. Rainfall and sea-surface temperature monitoring for malaria early warning in Botswana. *Am. J. Trop. Med. Hyg.* **73**, 214–221 (2005).
5. Greenwood, B. M., Bojang, K., Whitty, C. & Targett, G. A. T. Malaria. *Lancet* **365**, 1487–1498 (2005).
6. Worrall, E., Rietveld, A. & Delacollette, C. The burden of malaria epidemics and cost-effectiveness of interventions in epidemic situations in Africa. *Am. J. Trop. Med. Hyg.* **71** (Suppl. 2), 136–140 (2004).
7. WHO. *Malaria Early Warning Systems, Concepts, Indicators and Partners. A framework for Field Research in Africa*. 1–84, Report no. WHO/CDS/RBM/2001.32 (WHO/Roll Back Malaria/Technical Support Network for Prevention and Control of Malaria, Geneva, 2001); (http://mosquito.who.int/cmc_upload/0/000/014/807/mews2.pdf).
8. Palmer, T. N. *et al.* Development of a European ensemble system for seasonal to inter-annual prediction. *Bull. Am. Meteorol. Soc.* **85**, 853–872 (2004).
9. Palmer, T. N. *et al.* Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* **33**, 163–193 (2005).
10. Goddard, L. *et al.* Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.* **21**, 1111–1152 (2001).
11. Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. & Balmaseda, M. A. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature* **392**, 370–373 (1998).
12. Cane, M., Eshel, G. & Buckland, R. Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature. *Nature* **16**, 3059–3071 (1994).
13. Coelho, C. A. S., Pezzulli, S., Balmaseda, M., Doblas-Reyes, F. J. & Stephenson, D. B. Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Clim.* **17**, 1504–1516 (2004).
14. Jolliffe, I. T. & Stephenson, D. B. In *A Practitioner's Guide in Atmospheric Science* (eds Jolliffe, I. T. & Stephenson, D. B.) 1–12 (I.T. Wiley, Chichester, 2003).
15. Richardson, D. S. Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **126**, 649–667 (2000).
16. DaSilva, J. *et al.* Improving epidemic malaria planning, preparedness and response in southern Africa. *Malaria J.* **3**, doi:10.1186/1475-2875-3-37 (2004).
17. Brinkmann, U. & Brinkmann, A. Malaria and health in Africa—the present situation and epidemiologic trends. *Trop. Med. Parasitol.* **42**, 204–213 (1991).
18. Mason, S. J. & Tyson, P. D. In *Droughts Vol. 1 A Global Assessment* (ed. Wilhite, D. A.) 113–134 (Routledge, New York, 2000).
19. Xie, P. & Arkin, P. A. Global monthly precipitation from satellite-observed outgoing longwave radiation. *J. Clim.* **11**, 137–164 (1998).
20. Hagedorn, R., Doblas-Reyes, F. J. & Palmer, T. N. The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concepts. *Tellus A* **57**, 219–233 (2005).
21. Uppala, S. *et al.* ERA-40: ECMWF 45-year reanalysis of the global atmosphere and surface conditions 1957–2002. *ECMWF Newsl.* **101**, 2–21 (2004).
22. New, M., Hulme, M. & Jones, P. D. Representing twentieth century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Clim.* **12**, 829–856 (1999).
23. New, M., Hulme, M. & Jones, P. D. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **13**, 2217–2238 (2000).
24. Mason, S. J. & Graham, N. E. Areas beneath the relative operating characteristics (ROC) and levels (ROL) curves: statistical significance and interpretation. *Q. J. R. Meteorol. Soc.* **128**, 2145–2166 (2002).

Acknowledgements We thank the WHO-AFRO Southern Africa Inter-Country Malaria Team (SAMC) in Zimbabwe and the National Malaria Control Unit in Botswana for enabling this study. The work reported here is part of the EU-funded DEMETER and ENSEMBLES projects. It was additionally supported financially by the UK Department for International Development through the Malaria Knowledge Programme, Liverpool School of Tropical Medicine, UK, and by a cooperative agreement from the US National Oceanic and Atmospheric Administration. The authors acknowledge considerable technical support from ECMWF staff and consultants, in particular from the Seasonal Forecast Section. The views herein contained are those of the authors and do not necessarily reflect the views of WHO, DfID, EU, NOAA or any of their sub-agencies.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.N.P. (tim.palmer@ecmwf.int).

LETTERS

Episodic growth of the Gondwana supercontinent from hafnium and oxygen isotopes in zircon

A. I. S. Kemp¹, C. J. Hawkesworth¹, B. A. Paterson¹ & P. D. Kinny²

It is thought that continental crust existed as early as 150 million years after planetary accretion¹, but assessing the rates and processes of subsequent crustal growth requires linking the apparently contradictory information from the igneous and sedimentary rock records. For example, the striking global peaks in juvenile igneous activity 2.7, 1.9 and 1.2 Gyr ago imply rapid crustal generation in response to the emplacement of mantle 'super-plumes', rather than by the continuous process of subduction^{2–4}. Yet uncertainties persist over whether these age peaks are artefacts of selective preservation⁵, and over how to reconcile episodic crust formation with the smooth crustal evolution curves inferred from neodymium isotope variations of sedimentary rocks^{6,7}. Detrital zircons encapsulate a more representative record of igneous events than the exposed geology^{1,8,9} and their hafnium isotope ratios reflect the time since the source of the parental magmas separated from the mantle. These 'model' ages are only meaningful if the host magma lacked a mixed or sedimentary source component¹⁰, but the latter can be diagnosed by oxygen isotopes, which are strongly fractionated by rock-hydrosphere interactions. Here we report the first study that integrates hafnium and oxygen isotopes, all measured *in situ* on the same, precisely dated detrital zircon grains. The data reveal that crust generation in part of Gondwana was limited to major pulses at 1.9 and 3.3 Gyr ago, and that the zircons crystallized during repeated reworking of crust formed at these times. The implication is that the mechanisms of crust formation differed from those of crustal differentiation in ancient orogenic belts.

Palaeozoic sedimentary rocks that accumulated along the palaeo-Pacific margin of the former Gondwana supercontinent are ideal for testing punctuated crustal growth models. Their deposition postdates the major envisaged periods of rapid global crustal growth (2.7 and 1.9 Gyr ago; ref. 2), and their average crustal Sm/Nd ratios⁷ and uniformity of detrital zircon age populations^{11–13} testify to the efficient mixing of material eroded from large areas of continental crust. Detrital zircons in these rocks have ages that range to 3.5 Gyr, spanning most of Earth's history. The two Ordovician greywackes (sandstones containing >15% clay) used in this study are from the eastern Australian segment of Gondwana, and were metamorphosed in the greenschist facies during formation of the Ordovician to Devonian Lachlan fold belt. We also analysed the inherited (pre-magmatic) zircon cores of two cordierite-bearing granite plutons that intruded 430 Myr ago, and those of an equivalent volcanic rock (Supplementary Table 1). The granitic magmas were generated from broadly metasedimentary precursors at mid-crustal depths¹⁴, but their abundant inherited zircons have the same range of ages^{12,13,15} and isotope compositions as do detrital zircons in the exposed metasedimentary country rocks.

Approximately 100 zircons from each rock were mounted in resin, polished to expose their centres and imaged by cathodoluminescence.

Oscillatory or sector zonation is consistent with a magmatic paragenesis for the majority of these, although five irregularly banded grains with low Th/U ratios (<0.10) may represent metamorphic growth. The U–Pb, oxygen and Lu–Hf isotope compositions of each zircon were acquired by sequential, *in situ* analysis using three different instruments. The crystallization ages were first determined by SHRIMP (Sensitive High Resolution Ion MicroProbe) U–Pb isotope analysis at the Curtin University of Technology, Perth, under routine operating conditions¹. Concordant grains, or those that were <10% discordant for U–Pb isotopes, were then analysed for oxygen isotopes with a Cameca ims 1270 ion microprobe at the University of Edinburgh, the analysis site being directly adjacent to the SHRIMP pit. Last, Hf isotope ratios were measured from the same part of the zircon at the University of Bristol with a Finnigan Neptune multi-collector inductively coupled plasma-mass spectrometer (ICP-MS) and 193-nm ArF laser (Supplementary Methods).

All data are given in Supplementary Tables 2, 3 and 4. Oxygen isotope ratios are expressed in the standard $\delta^{18}\text{O}$ notation, signifying deviation of the measured $^{18}\text{O}/^{16}\text{O}$ value from Vienna standard mean ocean water (VSMOW) in parts per thousand. Zircons in equilibrium with pristine mantle-derived melts have $\delta^{18}\text{O}$ values of $5.3 \pm 0.3\text{‰}$ (ref. 16). Higher $\delta^{18}\text{O}$ values reflect a component of ^{18}O -enriched, supracrustal material in the magma from which the zircon precipitated, which could include sedimentary rock (10–40‰) or hydrothermally altered oceanic crust (10–20‰; ref. 17). Oxygen isotope ratios of igneous rocks are most sensitive to sedimentary input, as shown by the elevated $\delta^{18}\text{O}$ of metasedimentary-derived granites (9–15‰; ref. 18). Following ref. 19, we infer that zircons with $\delta^{18}\text{O}$ of less than 6.5‰ formed from melts that contained a minor to negligible sedimentary component. Zircons of this composition would crystallize from mafic melts with $\delta^{18}\text{O}$ values of <7‰, or silicic magmas with $\delta^{18}\text{O}$ below 8‰, assuming equilibrium melt/zircon fractionation²⁰. During this study we observed abrupt contrasts in $\delta^{18}\text{O}$ of nearly 6‰ between inherited zircon cores and their thin (<50 μm) magmatic overgrowths. This accords with the empirically established low diffusion rate of oxygen in zircon under high-temperature conditions²¹, and suggests that the measured $\delta^{18}\text{O}$ of inherited zircons approximates the primary value.

The $^{176}\text{Hf}/^{177}\text{Hf}$ data for detrital and inherited zircons with different $\delta^{18}\text{O}$ values are plotted as a function of their crystallization age in Fig. 1. Two striking features are evident. First, very few zircons have a Hf isotope composition that approaches that of the depleted mantle at the time of crystallization. The magmas from which most zircons formed were derived by melting pre-existing, rather than juvenile, crustal rocks. Second, zircons with $\delta^{18}\text{O} < 6.5\text{‰}$ define two conspicuous linear arrays that intersect the depleted mantle growth curve near 1.9 and 3.3 Gyr ago. The low $\delta^{18}\text{O}$ precludes a significant

¹Earth Sciences Department, University of Bristol, Bristol BS8 1RJ, UK. ²Tectonics Special Research Centre, Department of Applied Geology, Curtin University of Technology, Perth 6845, Australia.

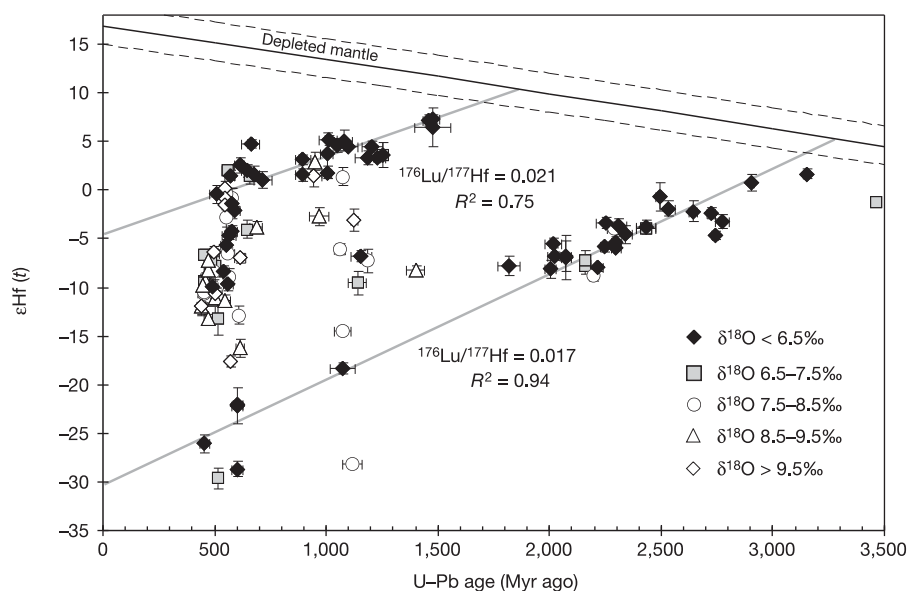


Figure 1 | Plot of ϵ_{Hf} versus inferred crystallization age for the inherited and detrital zircons analysed in this study, contoured for oxygen isotope composition. Error bars are indicated at 2 s.e.m., and the typical uncertainty of an oxygen isotope analysis is $\pm 0.4\text{‰}$ at 2 s.e.m. Note that the low $\delta^{18}\text{O}$ data point at 600 Myr ago and $\epsilon_{\text{Hf}} = -22$ comprises two almost identical analyses from separate zircons, and that the five inferred metamorphic zircons have $\delta^{18}\text{O} > 8.5\text{‰}$. The $^{207}\text{Pb}/^{206}\text{Pb}$ age is used for zircons exceeding 1.2 Gyr, and all others represent $^{206}\text{Pb}/^{238}\text{U}$ ages corrected for common lead by the ^{204}Pb method¹. Epsilon Hf values, representing deviation of the measured $^{176}\text{Hf}/^{177}\text{Hf}$ ratio from that of chondritic meteorites in parts per ten thousand, were calculated using the chondritic parameters of ref. 22 and $\lambda^{176}\text{Lu} = 1.867 \times 10^{-11} \text{ yr}^{-1}$ (ref. 23). The depleted mantle evolution curve (solid line) was extrapolated from the average modern-day values of $^{176}\text{Hf}/^{177}\text{Hf} = 0.28325$ and $^{176}\text{Lu}/^{177}\text{Hf} = 0.0334$, and is very similar to the curve delineated by the Hf

isotope composition of juvenile rocks through time²⁴. Dashed lines either side of this curve encompass most of the range shown by radiogenic modern mid-ocean ridge basalt (from ref. 31 and references therein). Regression lines (grey, with R^2 values indicated) through the two arrays defined by low $\delta^{18}\text{O}$ zircons correspond to the Hf isotope evolution of protoliths extracted from the mantle at either $1.87^{+0.19}_{-0.12}$ Gyr or 3.28 ± 0.10 Gyr, with $^{176}\text{Lu}/^{177}\text{Hf} = 0.021$ or 0.017 , respectively (depleted mantle intercepts are unchanged if ϵ_{Hf} values are calculated with the revised chondritic values of ref. 32). These $^{176}\text{Lu}/^{177}\text{Hf}$ ratios fall within the range of values shown by juvenile igneous rocks of all ages²⁴. The zircons that plot between the linear arrays presumably reflect melting of mixed-age igneous or metasedimentary precursors during crustal reworking. Regressions and errors on the slope of the low $\delta^{18}\text{O}$ zircon arrays were calculated using the Isoplot routines of K. Ludwig; available at <http://www.bgc.org/klprogramm.html>.

sedimentary source component for the magmas from which these zircons precipitated. We infer that each linear trend records the periodic, closed system re-melting of igneous material that was extracted from a depleted mantle reservoir at 1.9 or 3.3 Gyr ago. Supporting this interpretation, the slope of each low- $\delta^{18}\text{O}$ zircon array corresponds to the Hf isotope evolution of protoliths with $^{176}\text{Lu}/^{177}\text{Hf}$ ratios similar to those of juvenile crustal rocks²⁴ (Fig. 1). Remarkably, the newly added crustal material retained a broadly mantle-like O isotope composition despite enduring a protracted residence in the crust, in some cases for up to 2.7 Gyr before the thermal event that triggered zircon crystallization. The zircon record preserves no evidence for depleted mantle input at times other than 1.9 Gyr and 3.3 Gyr ago.

Equally conspicuous on Fig. 1 is that many zircons, particularly those with elevated (non-mantle-like) $\delta^{18}\text{O}$, have Hf isotope compositions that fall between the linear arrays of low $\delta^{18}\text{O}$ zircons. We contend that the sources to the magmas from which these zircons crystallized comprised mixtures of materials that originally separated from the mantle at 1.9 and 3.3 Gyr ago. This is an inevitable consequence of partial melting within heterogeneous continental crust, and accords with the lack of mixed compositions before the 1.9-Gyr-age crust-forming event. Mixing is accomplished by tectonic juxtaposition or by erosion and sedimentation in a supracrustal environment. Metamorphism and remelting of supracrustal protoliths explains the high $\delta^{18}\text{O}$ values of most of these zircons, especially those with $\delta^{18}\text{O} > 9\text{‰}$. The secular rise in $\delta^{18}\text{O}$ after 2.5 Gyr shown by the detrital and inherited zircons mirrors the trend defined by a global data set of magmatic zircons (Fig. 2), reaffirming the increased recycling of supracrustal materials by silicic magmas in the post-Archaean Earth²⁰.

Figure 1 reveals that crustal growth was fundamentally episodic in the segment of Gondwana sampled by zircons of this study, and confined to two major pulses at 1.9 and 3.3 Gyr ago. Juvenile rocks of these ages crop out in the Australian, Antarctic and southern African sectors of the former supercontinent²⁵. We note in particular the evidence for severe mantle depletion and attendant crust-formation 3.2–3.3 Gyr ago in the Kaapvaal craton, southern Africa²⁶. The 1.9-Gyr period also coincides with a global peak in mantle-derived magmatism³. Somewhat surprisingly, however, the 1.9-Gyr and 3.3-Gyr periods are not clearly recorded by zircon crystallization ages (Fig. 3). It is possible that the U–Pb isotope system in these zircons was reset by subsequent thermal events, although the selective obliteration of two zircon age populations seems unlikely. The preferred alternative is that crustal generation did not involve significant zircon crystallization. Instead, the zircons define prominent age peaks at 500–650 Myr ago and 0.9–1.2 Gyr ago that manifest Pan-African and Grenvillian orogenesis, respectively. These peaks register magmatic episodes related to crustal reworking and stabilization, rather than spikes in the rate of new continental crust generation. The Pan-African event is classically regarded as a time of crustal recycling, involving anatexis of metasedimentary precursors²⁷, as highlighted by the numerous Pan-African aged zircons with strongly elevated $\delta^{18}\text{O}$ values (Fig. 2). Many Grenvillian terranes are thought to represent juvenile crustal additions^{2,3,28}, but this study demonstrates that remelting of old and supracrustal rocks was regionally significant during this period. A similar conclusion has recently been reached for the Grenville province of North America²⁰.

In the light of these findings, crustal evolutionary models based on neodymium model ages of sedimentary rocks require revision. The Nd model age represents the average crustal residence time of the

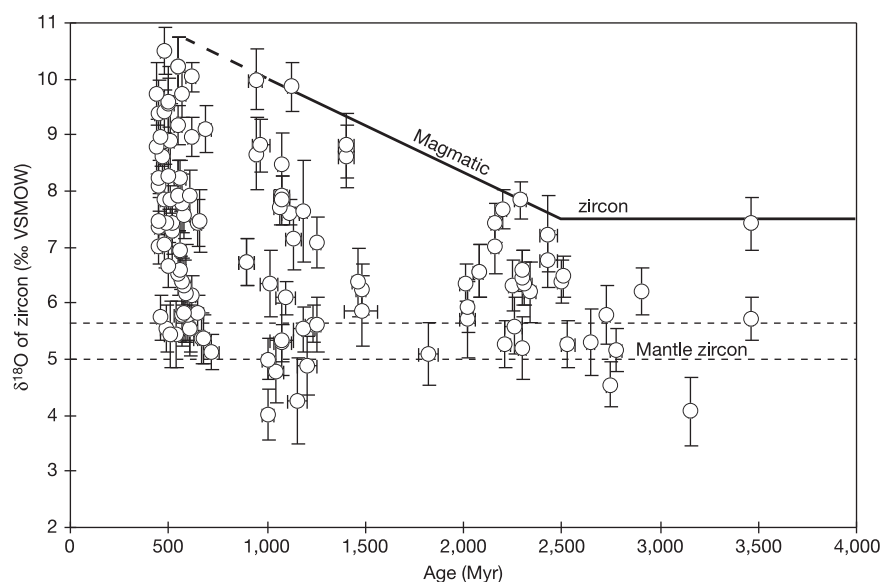


Figure 2 | Oxygen isotope composition of all detrital and inherited zircons analysed by this study as a function of crystallization age (error bars at 2 s.e.m.). The dashed horizontal lines bracket the $\delta^{18}\text{O}$ range of mantle zircon ($5.3 \pm 0.3\text{‰}$, ref. 16). The remarkable correspondence between the highest $\delta^{18}\text{O}$ values and the secular evolution in $\delta^{18}\text{O}$ exhibited by zircons

from a global compilation of 1,200 igneous rocks²⁰ (solid line) confirms that the zircons of this study were derived from a significant portion of continental crust. The rise in $\delta^{18}\text{O}$ recorded by the zircons testifies to the increasing incorporation of high $\delta^{18}\text{O}$ supracrustal materials by silicic magmas through time²⁰.

rocks in the sedimentary provenance. For Australian shales, the smooth decrease of Nd model age with stratigraphic age (Fig. 3) is attributed to the emplacement of mantle-derived igneous rocks⁷. Analogous curves are defined by the sedimentary rocks of Laurentia⁶. The implication of those studies is that new material has been continuously added to the continental mass throughout Earth's history, with the hyperbolic geometry of the curves reflecting a gradual reduction in this growth rate through time⁷. This interpretation now appears untenable for the provenance to the Gondwana metagreywackes. Alternatively, the decreasing Nd model ages may reflect the progressive contribution to the sedimentary column of

materials extracted from the mantle at 1.9 Gyr, perhaps owing to burial of the older rocks, or their return to the mantle.

The apparent decoupling between crust generation and zircon crystallization may mean that the processes controlling crustal growth and differentiation were intrinsically different. For this reason, zircon crystallization ages (which reflect thermal events and magmatic differentiation) can reveal little about the growth history of the continental crust, unless these are augmented by Lu–Hf and O isotope systematics. Conceivably, crust formation in the source area of the Gondwana metagreywackes involved the emplacement, perhaps by underplating, of zircon-free mafic rocks. The periodicity of this supports models that invoke crustal growth by the arrival of mantle plumes^{2–4}. The intracrustal differentiation and silicic magma genesis required to stabilize zircon may have been delayed by 100 to 400 million years and occurred during periods of intense crustal recycling involving supracrustal materials, but without detectable depleted mantle input. The unbroken (albeit peaked), distribution of zircon crystallization ages between the crustal growth episodes is consistent with a geologically continuous process like subduction. However, this scenario contrasts with modern subduction zones, where crust generation and differentiation are linked^{29,30}.

Received 7 August; accepted 29 November 2005.

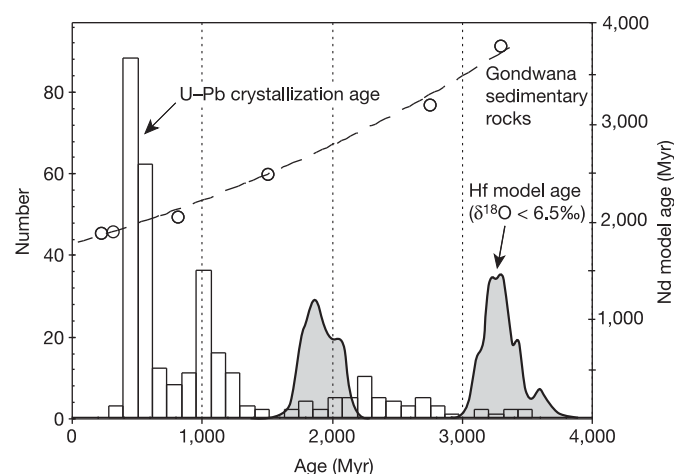


Figure 3 | Comparison between the crystallization ages of detrital and inherited zircons and the gaussian probability distribution of Hf model ages for zircons of the low $\delta^{18}\text{O}$ arrays. Peaks in these Hf model ages uniquely pinpoint times of juvenile crustal addition. All zircons are from southeastern Australia ($n = 318$) and incorporate data from refs 11–13. The dashed curve delineates the variation in Nd model ages of Gondwana sedimentary rocks through time (open circles, from ref. 7), as indicated on the right-hand-side y axis. Hafnium model ages were calculated using the $^{176}\text{Lu}/^{177}\text{Hf}$ ratios deduced from Fig. 1. 'Number' is the number of zircons with ages within the age groups indicated.

1. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
2. Stein, M. & Hofmann, A. W. Mantle plumes and episodic crustal growth. *Nature* **372**, 63–68 (1994).
3. Condie, K. C. Episodic continental growth and supercontinents: a mantle avalanche connection? *Earth Planet. Sci. Lett.* **163**, 97–108 (1998).
4. Albarède, F. The growth of continental crust. *Tectonophysics* **296**, 1–14 (1998).
5. Gurnis, M. & Davies, G. F. Apparent episodic crustal growth arising from a smoothly evolving mantle. *Geology* **14**, 396–399 (1986).
6. O'Nions, R. K., Hamilton, P. J. & Hooker, P. J. A Nd isotope investigation of sediments related to crustal development in the British Isles. *Earth Planet. Sci. Lett.* **63**, 229–240 (1983).
7. Allegre, C. J. & Rousseau, D. The growth of continents through geological time studied by Nd isotope analysis of shales. *Earth Planet. Sci. Lett.* **67**, 19–34 (1984).
8. Knudsen, T.-L., Griffin, W. L., Hartz, E. H., Andresen, A. & Jackson, S. E. In situ hafnium and lead isotope analysis of detrital zircons from the Devonian sedimentary basin of NE Greenland: a record of repeated crustal reworking. *Contrib. Mineral. Petrol.* **141**, 83–94 (2001).

9. Griffin, W. L., Belousova, E. A., Shee, S. R., Pearson, N. J. & O'Reilly, S. Y. Archaean crustal evolution in the northern Yilgarn Craton: U–Pb and Hf-isotope evidence from detrital zircons. *Precamb. Res.* **131**, 231–282 (2004).
10. Arndt, N. & Goldstein, S. L. Use and abuse of crust-formation ages. *Geology* **15**, 893–895 (1987).
11. Ireland, T. R., Flottman, T., Fanning, C. M., Gibson, G. M. & Preiss, W. V. Development of the early Palaeozoic Pacific margin of Gondwana from detrital zircon ages across the Delamerian Orogen. *Geology* **26**, 243–246 (1998).
12. Keay, S., Steele, D. & Compston, W. Identifying granite sources by SHRIMP U–Pb zircon geochronology: an application to the Lachlan Fold Belt. *Contrib. Mineral. Petrol.* **137**, 323–341 (1999).
13. Williams, I. S. Response of detrital zircon and monazite, and their U–Pb isotopic systems, to regional metamorphism and host rock partial melting, Cooma Complex, southeastern Australia. *Aust. J. Earth Sci.* **48**, 557–580 (2001).
14. Clemens, J. D. S-type granites- models and evidence. *Earth Sci. Rev.* **61**, 1–18 (2003).
15. Maas, R., Nicholls, I. A., Greig, A. & Nemchin, S. U–Pb zircon studies of mid-crustal metasedimentary enclaves from the S-type Deddick Granodiorite, Lachlan Fold Belt, SE Australia. *J. Petrol.* **42**, 1429–1448 (2001).
16. Valley, J. W. Oxygen isotopes in zircon. In *Zircon* (eds Hanchar, J. M. & Hoskin, P. W. O.) *Rev. Mineral.* **53**, 343–385 (2003).
17. Eiler, J. M. Oxygen isotope variations of basaltic lavas and upper mantle rocks. In *Stable Isotope Geochemistry* (eds Valley, J. W. & Cole, D. R.) *Rev. Mineral.* **43**, 319–364 (2001).
18. O'Neil, J. R. & Chappell, B. W. Oxygen and hydrogen isotope relations in the Berridale Batholith. *J. Geol. Soc. Lond.* **133**, 559–571 (1977).
19. Cavosie, A. J., Valley, J. W., Wilde, S. A. & E.I.M.F. Magmatic $\delta^{18}\text{O}$ in 4400–3900 detrital zircons: a record of the alteration and recycling of crust in the Early Archaean. *Earth Planet. Sci. Lett.* **235**, 663–681 (2005).
20. Valley, J. W. et al. 4.4 Billion years of crustal maturation: oxygen isotope ratios of magmatic zircon. *Contrib. Mineral. Petrol.* **150**, 561–580 (2005).
21. Peck, W. H., Valley, J. W. & Graham, C. M. Slow diffusion rates of O isotopes in igneous zircons from metamorphic rocks. *Am. Mineral.* **88**, 1003–1014 (2003).
22. Blichert-Toft, J. & Albarède, F. The Lu–Hf isotope geochemistry of chondrites and the evolution of the mantle-crust system. *Earth Planet. Sci. Lett.* **148**, 243–258 (1997).
23. Scherer, E., Münker, C. & Mezger, K. Calibration of the lutetium-hafnium clock. *Science* **293**, 683–687 (2001).
24. Vervoort, J. D. & Blichert-Toft, J. Evolution of the depleted mantle: Hf isotope evidence from juvenile rocks through time. *Geochim. Cosmochim. Acta* **63**, 533–566 (1999).
25. Goodwin, A. M. *Precambrian Geology* 1–666 (Academic Press, London, 1991).
26. Shirley, S. B. et al. Regional patterns in the paragenesis and age of inclusions in diamond, diamond composition, and the lithospheric seismic structure of Southern Africa. *Lithos* **71**, 243–258 (2003).
27. Harris, N. B. W., Hawkesworth, C. J. & Ries, A. C. Crustal evolution in northeast and east Africa from Nd model ages. *Nature* **309**, 773–776 (1984).
28. Patchett, P. J. & Ruiz, J. Nd isotopes and the origin of Grenville-aged rocks in Texas: implications for Proterozoic evolution of the United States mid-continent region. *J. Geol.* **97**, 685–695 (1989).
29. Tamura, Y. & Tasumi, Y. Remelting of an andesitic crust as a possible origin for rhyolitic magma in oceanic arcs: an example from the Izu-Bonin arc. *J. Petrol.* **43**, 1029–1047 (2002).
30. Vogel, T. A., Patino, L. C., Alvarado, G. E. & Gans, P. B. Silicic ignimbrites within the Costa Rican volcanic front: evidence for the formation of continental crust. *Earth Planet. Sci. Lett.* **226**, 149–159 (2004).
31. Chauvel, C. & Blichert-Toft, J. A hafnium isotope and trace element perspective on melting of the depleted mantle. *Earth Planet. Sci. Lett.* **190**, 137–151 (2001).
32. Patchett, P. J., Vervoort, J. D., Söderlund, U. & Salters, V. J. M. Lu–Hf and Sm–Nd isotopic systematics in chondrites and their constraints on the Lu–Hf properties of the Earth. *Earth Planet. Sci. Lett.* **222**, 29–41 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to J. Craven and C. Graham (Ion Microprobe Facility, University of Edinburgh) for advice and guidance in obtaining the oxygen isotope data, G. Foster, C. Coath and A. Scherstén for assistance in the Bristol laboratory, and T. Elliot for comments on an earlier draft. The manuscript benefited from reviews by J. Valley and U. Söderlund.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.K. (tony.kemp@geo.sc.niigata-u.ac.jp).

LETTERS

Early evolution of the venom system in lizards and snakes

Bryan G. Fry^{1,2}, Nicolas Vidal^{3,4}, Janette A. Norman², Freek J. Vonk⁵, Holger Scheib^{6,7}, S. F. Ryan Ramjan¹, Sanjaya Kuruppu⁸, Kim Fung⁹, S. Blair Hedges³, Michael K. Richardson⁵, Wayne. C. Hodgson⁸, Vera Ignjatovic^{10,11}, Robyn Summerhayes^{10,11} & Elazar Kochva¹²

Among extant reptiles only two lineages are known to have evolved venom delivery systems, the advanced snakes and helodermatid lizards (*Gila Monster* and *Beaded Lizard*)¹. Evolution of the venom system is thought to underlie the impressive radiation of the advanced snakes (2,500 of 3,000 snake species)^{2–5}. In contrast, the lizard venom system is thought to be restricted to just two species and to have evolved independently from the snake venom system¹. Here we report the presence of venom toxins in two additional lizard lineages (Monitor Lizards and Iguania) and show that all lineages possessing toxin-secreting oral glands form a clade, demonstrating a single early origin of the venom system in lizards and snakes. Construction of gland complementary-DNA libraries and phylogenetic analysis of transcripts revealed that nine toxin types are shared between lizards and snakes. Toxinological analyses of venom components from the Lace Monitor *Varanus varius* showed potent effects on blood pressure and clotting ability, bioactivities associated with a rapid loss of consciousness and extensive bleeding in prey. The iguanian lizard *Pogona barbata* retains characteristics of the ancestral venom system, namely serial, lobular non-compound venom-secreting glands on both the upper and lower jaws, whereas the advanced snakes and anguimorph lizards (including Monitor Lizards, *Gila Monster* and *Beaded Lizard*) have more derived venom systems characterized by the loss of the mandibular (lower) or maxillary (upper) glands. Demonstration that the snakes, iguanians and anguimorphs form a single clade provides overwhelming support for a single, early origin of the venom system in lizards and snakes. These results provide new insights into the evolution of the venom system in squamate reptiles and open new avenues for biomedical research and drug design using hitherto unexplored venom proteins.

In helodermatid lizards, venom is made by a gland on the lower jaw from which ducts lead onto grooved teeth along the length of the mandible¹. In contrast, snake venom is produced by specialized glands in the upper jaw^{1,6–12} and is a shared derived trait of the advanced snakes^{2–5,13}. To investigate the evolution of venomous function in squamates we first obtained sequence data from five nuclear protein-coding genes representing major squamate lineages. Our phylogenetic analyses show that the closest relatives of snakes are the anguimorph (which include the venomous helodermatids) and iguanian lizards (Fig. 1, and Supplementary Fig. 1). These results

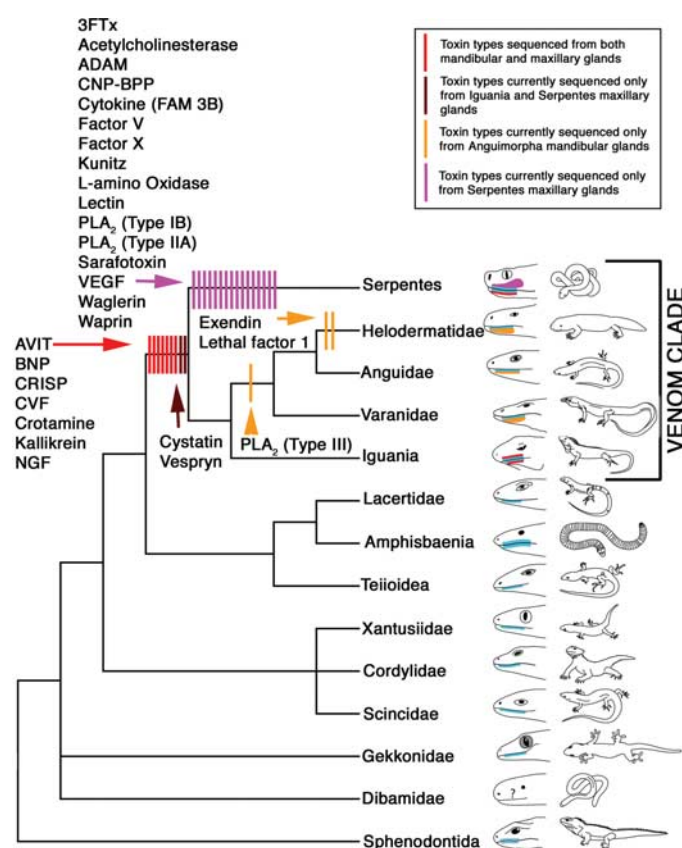


Figure 1 | Relative glandular development and timing of toxin recruitment events mapped over the squamate reptile phylogeny. Mucus-secreting glands are coloured blue; the ancestral form of the protein-secreting gland (serial, lobular and non-compound) red; the complex, derived form of the upper snake-venom gland (compound, encapsulated and with a lumen) fuchsia, and the complex, derived form of the anguimorph mandibular venom gland (compound, encapsulated and with a lumen) orange. 3FTx, three-finger toxins; ADAM, a disintegrin and metalloproteinase; CNP-BPP, C-type natriuretic peptide-bradykinin-potentiating peptide; CVP, cobra venom factor; NGF, nerve growth factor; VEGF, vascular endothelial growth factor.

¹Australian Venom Research Unit, Level 8, School of Medicine, University of Melbourne, Parkville, Victoria 3010, Australia. ²Population and Evolutionary Genetics Unit, Museum Victoria, GPO Box 666E, Melbourne, Victoria 3001, Australia. ³Department of Biology and Astrobiology Research Center, 208 Mueller Lab, Pennsylvania State University, University Park, Pennsylvania 16802-5301, USA. ⁴UMS 602, Taxonomie et collections, Reptiles-Amphibiens, Département Systématique et Évolution, Muséum National d'Histoire Naturelle, 25 Rue Cuvier, Paris 75005, France. ⁵Institute of Biology, Leiden University, Kaiserstraat 63, PO Box 9516, 2300 RA, Leiden, The Netherlands. ⁶Department of Structural Biology and Bioinformatics, University of Geneva and Swiss Institute of Bioinformatics, Centre Médical Universitaire, 1 Rue Michel-Servet, 1211 Geneva 4, Switzerland. ⁷SBC Lab AG, Seebühlstrasse 26, 8185 Winkel, Switzerland. ⁸Monash Venom Group, Department of Pharmacology, Monash University, Clayton, Victoria 3800, Australia. ⁹Molecular and Health Technologies, CSIRO, 343 Royal Parade, Parkville, Victoria 3010, Australia. ¹⁰Department of Pathology, University of Melbourne, Parkville, Victoria 3010, Australia. ¹¹Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. ¹²Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel.

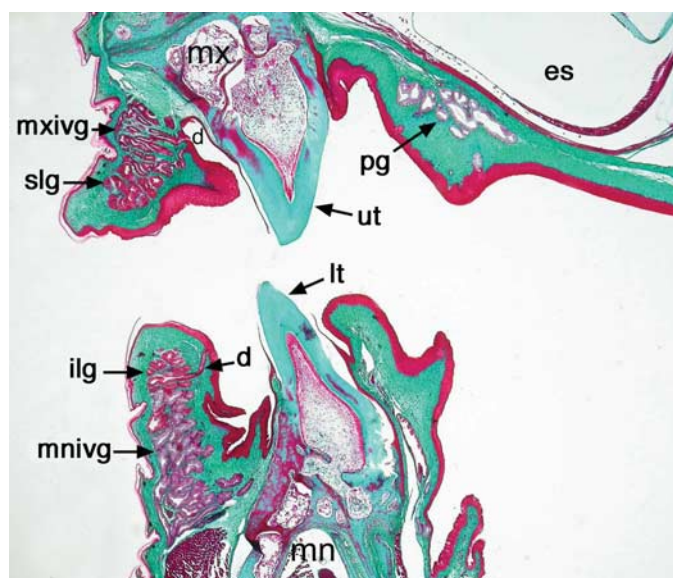


Figure 2 | Transverse section of *Pogona barbata* (Eastern Bearded Dragon) head to show relative arrangement of glands. Stain: Masson's trichrome. Original magnification $\times 40$. Abbreviations: d, duct; es, eye socket; ilg, infralabial gland; lt, lower tooth; mn, mandible; mnivg, mandibular incipient venom gland; mx, maxillary; mxivg, maxillary incipient venom gland; pg, palatine gland; slg, supralabial gland; ut, upper tooth.

represent a major paradigm shift in the understanding of squamate evolution. The three lineages were previously considered as part of a large, unresolved polytomy that also included amphisbaenian, lacertid and teioid lizards^{14,15}. We subsequently mapped the structure of protein-secreting oral glands (mandibular and maxillary) over our revised squamate phylogeny (Fig. 1). The anguimorphs,

iguanians and snakes, which form a well-resolved clade, are shown to be the only lineages possessing protein-secreting mandibular and/or maxillary glands. The presence of a protein-secreting gland is therefore a shared derived trait of this entire clade. The basal condition of a serial, lobular and non-compound protein-secreting gland, present in both mandibular and maxillary regions, is retained in the iguanian lizards (Fig. 2). The restriction of protein-secreting function to the maxillary (advanced snakes) or mandibular (anguimorphs) glands represents highly derived conditions. In the respective regions, the snakes (maxillary) and anguimorphs (mandibular) have independently evolved complex, compound venom glands with encapsulation and lumen formation for the storage of liquid venom for ready delivery^{1,16,17}. Some snakes (for example *Natrix*) still express proteins in their serial, lobular, non-compound mandibular glands, whereas the anguimorphs have lost the maxillary glands entirely¹⁸.

To explain the distribution, recruitment¹⁹ and molecular evolution²⁰ of venom proteins among them, representatives of the anguimorph/iguanian/serpent clade, spanning a wide ecological breadth, were also investigated for the secretion of toxins through the construction of cDNA libraries. Mandibular gland cDNA libraries were constructed for the Eastern Bearded Dragon (an iguanian) and four varanids (anguimorphs); maxillary-gland cDNA libraries were also constructed for the Eastern Bearded Dragon. To provide greater coverage of venom evolution, maxillary-gland cDNA libraries were also constructed from seven advanced snake families. Transcripts coding for previously characterized lizard or snake toxin types were identified in all gland cDNA libraries and we report the presence of venom toxins in lizards (other than *Heloderma*). Nine toxin types were recovered from both lizard and snake cDNA libraries (AVIT, B-type natriuretic peptide (BNP), CRISP, cobra venom factor, crotoamine, cystatin, kallikrein, nerve growth factor and vespryn), being secreted from mandibular and maxillary glands. Bayesian phylogenetic analyses^{13,19} of these nine toxin types resulted in the monophyly of each venom toxin to the

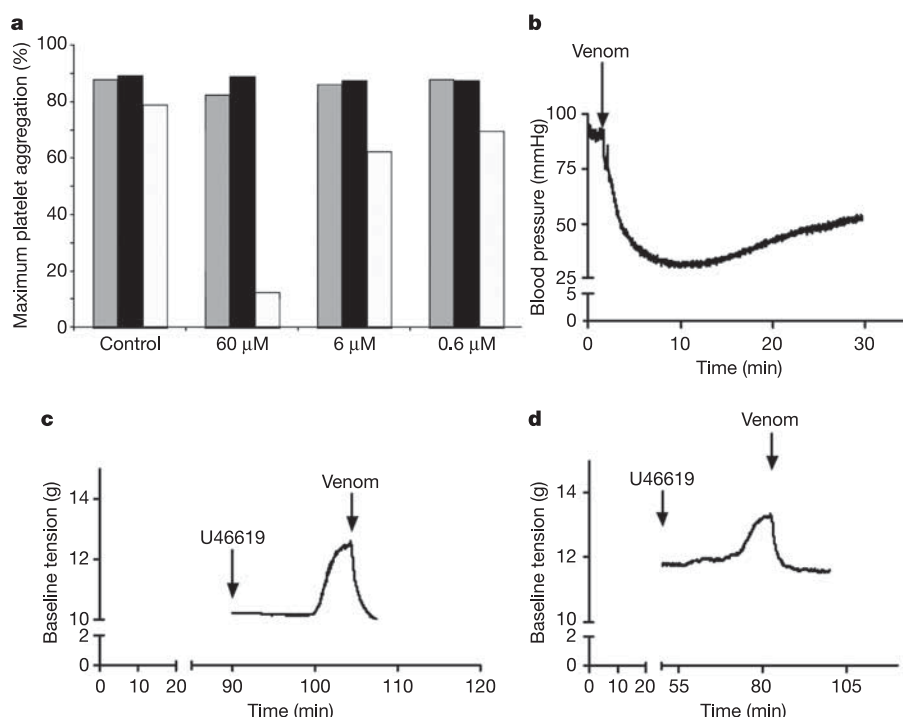


Figure 3 | Bioactivity of *V. varius* (Lace Monitor) venom. **a**, The effect of different molar concentrations of purified type III PLA₂ toxin DQ139930 on platelet aggregation. The control was saline. Grey bars, 2 μ M ADP; black bars, 30 μ M adrenaline; white bars, 5 μ M adrenaline. **b**, Effect of intravenous

injection of crude venom (1 mg kg⁻¹) on blood pressure in the anaesthetized rat. **c**, **d**, Effect of crude venom (200 μ g ml⁻¹) on U46619 precontracted endothelium-intact (**c**) and endothelium-denuded (**d**) aortic rings. $n = 3$; single traces are shown.

exclusion of related non-venom proteins. This pattern strongly supports single early recruitment events¹³ for each toxin type before the separation of snakes, iguanians and anguimorphs (Supplementary Figs 2–10). An additional toxin, type III phospholipase A₂ (PLA₂), previously characterized only from *Heloderma* venoms²¹ (Gila Monster and Beaded Lizard), was identified in varanid mandibular glands.

The mapping of toxin types, revealed in this study as being secreted in both mandibular and maxillary glands, over the revised squamate phylogeny provides additional insights into the evolution of the reptile venom chemical arsenal (Fig. 1). Most striking is the proposed complexity of the venom secretions in the common ancestor of venomous lizards and snakes with nine toxin types present. Seven of these were previously known only from snake venoms, including one toxin type (crotonamine), sequenced from the Eastern Bearded Dragon mandibular and maxillary glands, previously characterized only from rattlesnake venoms. The nine shared venom toxins isolated all possess previously well-characterized activities, including hypolocomotion, hypotension, hypothermia, immunomodulatory effects, intestinal cramping, myonecrosis, paralysis of peripheral smooth muscle unregulated activation of the complement cascade, and hyperalgesia¹⁹. The type III PLA₂ toxins from *Heloderma* venoms have been shown to block platelet aggregation²¹. Some of these toxins have been shown to have potent systemic effects, such as the profound hypotension produced by kallikrein and natriuretic toxins¹⁹ leading to rapid loss of consciousness, or coagulation disorders such as prolonged bleeding as a consequence of the *Heloderma* type III PLA₂ toxins²¹. Some toxins exert effects that, although non-lethal, may aid in the rapid incapacitation of prey items or potential predators, such as the markedly increased sensitivity to pain (hyperalgesia) and strong cramping produced by the AVIT toxins¹⁹.

Varanid venom was revealed by toxinological analyses to be as complex and potent as previously analysed reptile venoms^{5,22}. Liquid chromatography/mass spectrometry⁵ (LC/MS) showed the varanid

secretions to be rich in proteins with molecular masses consistent with the following toxin types sequenced from varanid cDNA libraries: natriuretic (2–4 kDa), type III PLA₂ (about 15 kDa), CRISP (23–25 kDa) and kallikrein (23–25 kDa) (Supplementary Fig. 11). Haematological assays of varanid type III PLA₂ toxin (DQ139930) purified by reverse-phase high-performance liquid chromatography²³ showed inhibition of platelet aggregation (Fig. 3a). Consistent with the same bioactivity as *Heloderma* type III PLA₂ (ref. 21) was the preservation of cysteines and cysteine spacing as well as the conservation of functional residues (Supplementary Fig. 12A). As cDNA sequencing and LC/MS analysis indicated a high concentration of kallikrein and BNP-type natriuretic toxins in the varanid secretions, additional assays investigated hypotension-inducing bioactivity. Intravenous injections of crude *Varanus varius* mandibular secretion to anaesthetized rats rapidly produced a sharp drop in blood pressure (Fig. 3b) and specific analyses with precontracted rat aortic rings²³ demonstrated the natriuretic peptide action of relaxation of aortic smooth muscle (Fig. 3c, d). Consistent with the preserved bioactivity of the varanid BNP-type natriuretic toxins, sequence analysis and molecular modelling revealed the retention of residues necessary for natriuretic activity²³ (Fig. 4, and Supplementary Fig. 10). The varanid forms of the kallikrein toxins also show conservation of functional residues and cysteine spacing (Supplementary Fig. 12B). In the CRISP toxins, the varanid forms all have the loop I doublet (KR) that has been proposed to be an essential part of the blockage of cyclic-nucleotide-gated calcium channels (Supplementary Fig. 13). Most varanid CRISP isoforms also have the loop I motif (EXXF) thought to contribute to the inhibition of smooth muscle contraction through the blockage of L-type Ca²⁺ channels (Supplementary Fig. 13). Other toxin types vary to differing degrees in the relative conservation of molecular characteristics.

The combined cDNA, LC/MS, molecular modelling and pharmacological results are consistent with effects reported for varanid bites,

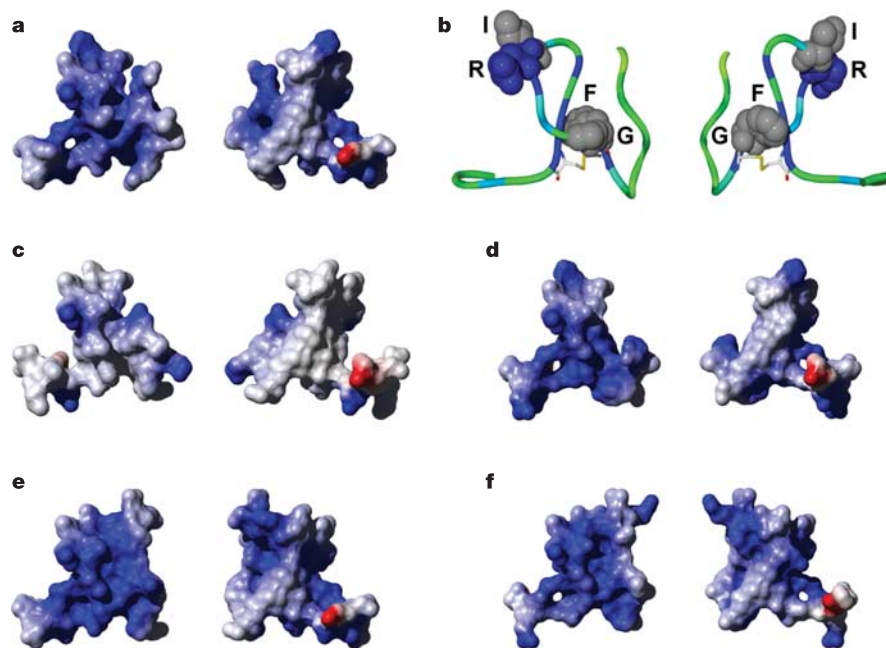


Figure 4 | Comparative modelling of representative natriuretic peptides.

a, GNP-1 (DQ139927) from *V. varius* (Lace Monitor); **c**, TNP-c (P83230) from *Oxyuranus microlepidotus* (Inland Taipan); **d**, DNP (P28374) from *Dendroaspis angusticeps* (Eastern Green Mamba); **e**, Lebetin (Q7LZ09) from *Macrovipera lebetina* (Elephant Snake); **f**, BNP from the rat (P13205) brain and atria. Blue surface areas indicate positive charges, red areas show negative charges. Model pairs show the sides of the protein rotated by 180°.

b, GNP in ribbon representation coloured by alignment diversity²⁹.

Conserved positions are in blue, brighter colours indicate increasing degree of variation. Functional residues²³ are CPK (Corey–Pauling–Koltun) coloured by amino-acid type²⁹. Hydrophobic residues are in grey, arginine in blue. The conserved cysteines are shown as sticks. To minimize confusion, all sequences are referred to by their SWISS-PROT accession numbers (<http://www.expasy.org/cgi-bin/sprot-search-ful>).

which include severe pain, breathing difficulties, skeletal muscle weakness and tachycardia²⁴. One of the authors (B.G.F.) has also acted as consultant on three varanid bites by captive bred specimens (*Varanus komodoensis* (Komodo Dragon), *V. scalaris* (Spotted Tree Monitor) and *V. varius* (Lace Monitor)), each of which resulted in rapid swelling (noticeable within minutes), dizziness, localized disruption of blood clotting and shooting pain extending from the affected digit up to the elbow, with some symptoms lasting for several hours. The rapidity and pathology are consistent with bioactive secretions rather than bacterial infection. In addition, varanid venom also has been shown to have the ability to rapidly paralyse small animals such as birds²⁵.

As well as providing evidence about the role of bioactive secretions in the effects produced by varanid bites, the complex and bioactive secretions present in 'non-venomous' lizards forces a fundamental rethinking of the very concept of 'non-venomous' reptile. The evolution of venomous function is considered to be a key innovation driving ecological diversification in advanced snakes^{2–5}. Our results indicate that the single origin of venom in squamate reptiles might also have been a key factor in the adaptive radiation and subsequent ecological success of lizard lineages such as anguimorphs and iguanians; the well-supported anguimorph/iguanaian/serpent clade represents about 4,600 out of about 7,900 extant squamate species, or 58% of the total squamate species diversity. There is also palaeontological evidence for the presence of venom delivery systems in some Upper Cretaceous anguimorphs and snakes^{8,26,27}. Because fossil data indicate that the clade containing anguimorphs, iguanians and snakes dates from Late Triassic/Early Jurassic²⁸, we infer that the venomous function arose once in squamate evolution, at about 200 Myr ago. This considerably revises previous estimates of about 100 Myr ago based on the assumed independent origin of venomous function in snakes and lizards.

Additional work aimed to investigate this special area of reptile evolution would include investigating the mandibular or maxillary secretions of all main lineages belonging to the toxin-secreting clade. Studies of additional transcriptomes may reveal earlier recruitment of a particular toxin type, such as those currently sequenced only from snake venoms for example, or may discover previously unrecognized toxin types. Further pharmacological investigations may shed more light upon the bioactivities and the relative use in defence, in prey capture or in predigestion. Because increased complexity of the venom gland was shown in this study to be linked to additional toxin recruitment events (Fig. 1), future work should include an exploration of the relationship between glandular complexity and the relative toxicity of the venom. The new lizard venom toxins exhibit considerable sequence diversity consistent with the birth-and-death mode of protein evolution that has given rise to a wide diversity of bioactivities in snake venoms²⁰. These molecules represent a tremendous hitherto unexplored resource not only for understanding reptile evolution but also for use in drug design and development.

METHODS

Toxin molecular evolution. Specimen collection localities: *Azemioops feae* (Fea's Viper), Guizhou, China; *Enhydryis polylepis* (Macleay's Water Snake), Darwin, Northern Territory, Australia; *Dispholidus typus* (Boomslang), Uganda; *Oxyuranus microlepidotus* (Inland Taipan), progeny of captive specimens from Goydnor's Lagoon, South Australia; *Pogona barbata* (Eastern Bearded Dragon), progeny of Brisbane, Queensland, Australia locality captive specimens; *Telescopus dhara* (Egyptian Catsnake), Egypt; *Trimorphodon biscutatus* (Lyre Snake), Arizona, USA; *Liophis poecilogyrus* (Gold-bellied Snake), Paraguay; *Leioheterodon madagascariensis* (Madagascar Giant Hognosed Snake), Madagascar; *Philodryas olfersii* (Argentine Racer), Argentina; *Rhabdophis tigrinus* (Tiger Keelback), Hunan, China; *Varanus acanthurus* (Spiny-tailed Monitor), progeny of captive specimens collected from Newman, Western Australia; *Varanus mitchelli* (Mitchell's Water Monitor), Kununurra, Western Australia; *Varanus panoptes rubidus* (Desert Spotted Goanna), Sandstone, Western Australia; *Varanus varius* (Lace Monitor), Mallacoota, Victoria, Australia.

RNA extraction and construction of cDNA library: these steps were undertaken with the Qiagen RNeasy and Oligotex messenger RNA kits and the Creator SMART cDNA Library Construction Kit from BD Biosciences. Full details are available in Supplementary Information.

Histology. Tissue was dissected from animals after killing, then fixed in Bouin's fluid and decalcified overnight in acid alcohol (5% HCl in 70% ethanol). The tissues were dehydrated in graded ethanols, cleared in Histoclear and embedded in paraffin. Sections were cut to 10 µm thickness and stained with Masson's trichrome (Goldner's modification), Alcian blue at pH 1.0 and 2.4 alone or counterstained with haematoxylin–eosin or periodic acid Schiff in accordance with standard techniques.

Molecular modelling. Three-dimensional models were generated by aligning target sequences to the 1Q01 template with SPDBV²⁹. Sequence-to-structure alignments were sent to the Swiss-Model server. For the resulting models a van der Waals surface was calculated in MolMol³⁰. Surfaces were coloured by a 'simplecharge' potential as calculated in MolMol.

Pharmacology. Male rats were anaesthetized with sodium pentobarbitone (60–100 mg kg^{−1}, intraperitoneally). Venom or vehicle (namely 0.1% BSA) was administered through the jugular-vein cannula. Blood pressure was measured with a Gould (P23) pressure transducer attached to a carotid artery cannula, and recorded on a MacLab system.

Platelet aggregometry. Blood samples were collected from normal, healthy adult volunteers ($n = 2$) who had not taken any medication during the week before collection. Whole-blood samples were mixed 9:1 with 0.106 M sodium citrate. Citrated blood samples were centrifuged for 10 min at 100 g at 20 °C. The supernatant platelet-rich plasma (PRP) was removed and rested at room temperature for 30 min before assay. Platelet-poor plasma (PPP) for each volunteer was also prepared (by centrifugation for 20 min at 3,500 g) for platelet aggregometry. Platelet aggregation was measured with the Aggram platelet aggregometer and Hemoram software (Helena Laboratories). PRP aliquots (225 µl) were incubated in glass cuvettes for 2 min at 37 °C. Purified *Varanus varius* PLA₂ toxin DQ139930 (60, 6 and 0.6 µM) was added to the PRP and incubated at 37 °C for 3 min before the addition of agonist (2 µM ADP, or 5 or 30 µM adrenaline), to induce platelet aggregation, or sterile saline as a control. Platelet aggregation was monitored for 10 min. The degree of maximum platelet aggregation was assessed by measuring the optical transmission of light, zeroed with the appropriate PPP for each patient sample, through the PRP samples and compared with the controls (PRP samples assayed without the addition of toxin).

Squamate reptile phylogeny. Recent studies^{14,15} included representatives of all major squamate lineages and identified a clade comprising the following five lineages: first, Scincoidea (Scincidae, Xantusiidae and Cordylidae); second, Teiioidea (Teiidae and Gymnophthalmidae), Lacertidae and Amphisbaenia (Rhineuridae, Bipedidae, Trogonophidae and Amphisbaenidae); third, Iguania (Iguanidae, Agamidae and Chamaeleontidae); fourth, Anguimorpha (Varanidae, Helodermatidae, Anguinae, *Shinisaurus* and *Xenosaurus*); and fifth, Serpentes. The venomous squamates (advanced snakes and helodermatid lizards) are included in this large clade, whereas two other families of squamates (Gekkonidae and Dibamidae) are excluded^{14,15}. We focused on this clade and sampled five nuclear protein-coding genes, including two genes (*C-mos* and *RAG1*) used in other studies and three genes (*RAG2*, *R35* and *HOXA13*) not used previously to clarify squamate phylogeny. Phylogenies were built with probabilistic approaches (maximum-likelihood (ML) and bayesian methods of inference). Because separate analyses showed no significant topological incongruence, we performed combined analyses, which are considered to be our best estimates of phylogeny. Scincoidea was used as the outgroup because it was shown to be the most basal lineage of the clade^{14,15}. The bayesian and ML trees obtained were identical and showed significant support for most nodes (see Supplementary Information). In particular, a clade that includes Serpentes, Iguania and Anguimorpha was resolved (bayesian posterior probability 100%; ML bootstrap value 99%). In turn, we found that the closest relative of this clade is one comprising Teiioidea, Lacertidae and Amphisbaenia. Full details are available in Supplementary Information.

Received 13 July; accepted 17 October 2005.

Published online 16 November 2005.

- Kochva, E. In *Biology of the Reptilia* Vol 8 (eds Gans, S. K. & Gans, C.) 43–162 (Academic, London, 1978).
- Vidal, N. Colubroid systematics: evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. *J. Toxicol. Toxin Rev.* 21, 21–41 (2002).
- Vidal, N. & Hedges, S. B. Higher-level relationships of caenophidian snakes inferred from four nuclear and mitochondrial genes. *C. R. Biol.* 325, 987–995 (2002).

4. Fry, B. G. *et al.* Isolation of a neurotoxin (alpha-cobrotoxin) from a 'non-venomous' colubrid: evidence for early origin of venom in snakes. *J. Mol. Evol.* **57**, 446–452 (2003).
5. Fry, B. G. *et al.* LC/MS (liquid chromatography, mass spectrometry) analysis of Colubroidea snake venoms: evolutionary and toxicological implications. *Rapid Commun. Mass Spectrom.* **17**, 2047–2062 (2003).
6. Kochva, E. Development of the venom gland and trigeminal muscles in *Vipera palaestinae*. *Acta Anat.* **52**, 49–89 (1963).
7. Kochva, E. The development of the venom gland in the opisthoglyph snake *Telescopus fallax* with remarks on *Thamnophis sirtalis* (Colubridae, Reptilia). *Copeia* **2**, 147–154 (1965).
8. Kochva, E. The origin of snakes and evolution of the venom apparatus. *Toxicon* **25**, 65–106 (1987).
9. Kochva, E. *Atractaspis* (Serpentes, Atractaspididae) the Burrowing Asp; a multidisciplinary minireview. *Bull. Nat. Hist. Mus. Lond. Zool.* **68**, 91–99 (2002).
10. Underwood, G. & Kochva, E. On the affinities of the burrowing asps *Atractaspis* (Serpentes: Atractaspididae). *Zool. J. Linn. Soc.* **107**, 3–64 (1993).
11. Underwood, G. in *Venomous Snakes: Ecology, Evolution and Snakebite* (eds Thorpe, R. S., Wüster, W. & Malhotra, A.) 1–13 (Symp. Zool. Soc. Lond. no. 70, Clarendon, Oxford, 1997).
12. Jackson, K. The evolution of venom-delivery systems in snakes. *Zool. J. Linn. Soc.* **137**, 337–354 (2003).
13. Fry, B. G. & Wüster, W. Assembling an arsenal: Origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol. Biol. Evol.* **21**, 870–883 (2004).
14. Vidal, N. & Hedges, S. B. Molecular evidence for a terrestrial origin of snakes. *Proc. R. Soc. Lond. B Suppl.* **271**, S226–S229 (2004).
15. Townsend, T. M., Larson, A., Louis, E. & Macey, J. R. Molecular phylogenetics of Squamata: The position of snakes, Amphisbaenians, and Dibamids, and the root of the squamate tree. *Syst. Biol.* **53**, 735–757 (2004).
16. Gabe, M. & Saint Girons, H. Données histologiques sur les glandes salivaires des lépidosauriens. *Mém. Mus. Natl. Hist. Nat. Paris* **58**, 1–118 (1969).
17. Gabe, M. & Saint Girons, H. in *Toxins of Animal and Plant Origin* (eds de Vries, A. & Kochva, E.) 65–68 (Gordon & Breach, London, 1971).
18. Gygas, P. Entwicklung, Bau und Funktion der Giftdrüse (Duvernoy's gland) von *Natrix tessellata*. *Acta Trop. Zool.* **28**, 225–274 (1971).
19. Fry, B. G. From genome to 'venome': Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* **15**, 403–420 (2005).
20. Fry, B. G. *et al.* Molecular evolution of elapid snake venom three finger toxins. *J. Mol. Evol.* **57**, 110–129 (2003).
21. Huang, T. F. & Chiang, H. S. Effect on human platelet aggregation of phospholipase A2 purified from *Heloderma horridum* (beaded lizard) venom. *Biochim. Biophys. Acta* **1211**, 61–68 (1994).
22. Fry, B. G. *et al.* Electrospray liquid chromatography/mass spectrometry fingerprinting of *Acanthophis* (death adder) venoms: taxonomic and toxicological implications. *Rapid Commun. Mass Spectrom.* **16**, 600–608 (2002).
23. Fry, B. G. *et al.* Novel natriuretic peptides from the venom of the inland taipan (*Oxyuranus microlepidotus*): Isolation, chemical and biological characterization. *Biochem. Biophys. Res. Commun.* **327**, 1011–1015 (2005).
24. Sopiev, O., Makeev, B. M., Kudryavtsev, S. B. & Makarov, A. N. A case of intoxication by a bite of the gray monitor (*Varanus griseus*). *Izv. Akad. Nauk Turkm. SSR. Ser. Biol. Nauk* **87**, 78 (1987).
25. Gorelov, Y. U. K. About the toxicity of the saliva of the gray monitor. *Izv. Akad. Nauk Turkm. SSR. Ser. Biol. Nauk* **71**, 74 (1971).
26. Pregill, G. K., Gauthier, J. A. & Greene, H. W. The evolution of Helodermatid squamates, with description of a new taxon and an overview of Varanoidea. *Trans. San Diego Soc. Nat. Hist.* **21**, 167–202 (1986).
27. Norell, M. A., McKenna, M. C. & Novacek, M. J. *Estesia mongoliensis*, a new fossil varanoid from the Late Cretaceous Barun Goyot Formation of Mongolia. *Am. Mus. Novit.* **3045**, 1–24 (1992).
28. Evans, S. E. At the feet of the dinosaurs: the origin, evolution and early diversification of squamate reptiles (Lepidosauria: Diapsida). *Biol. Rev. Camb.* **78**, 513–551 (2003).
29. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
30. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**(1M), 51–55 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the following persons and institutions who helped us or contributed tissue samples used in this study: A. Fry, Alice Springs Reptile Centre, Australian Reptile Park, M. A. G. de Bakker, R. L. Bezy, B. Branch, J. Campbell, N. Clemann, C. Clemente, C. Cicero, K. Daoues, A. S. Delmas, B. Demeter, J. Haberfield, A. Hassanin, Healesville Sanctuary, M. Hird, Louisiana State University Museum of Zoology, P. Moler, T. Moncuit, P. Moret, National Museum of Natural History Naturalis Leiden (J. W. Arntzen), T. Pappenfus, J.-C. Rage, C. Skliris, J. Smith, S. Sweet, Ultimate Reptiles (South Australia), University of California Museum of Vertebrate Zoology (Berkeley), J. Walker, R. Waters, J. Weigel and B. Wilson. We also thank A. Webb and T. Purcell for providing HPLC access; N. Williamson for help with preliminary mass spectrometry characterization; E. V. Grishin for help in obtaining the references in Russian; S. Edwards for comments; and T. van Wagner and V. Wexler for artwork. This work was funded by the Service de Systématique moléculaire of the Muséum National d'Histoire Naturelle, Institut de Systématique (N.V.) and by grants from the Australian Academy of Science (B.G.F.), Australian Geographic Society (B.G.F.), Australia & Pacific Science Foundation (B.G.F.), Australian Research Council (B.G.F.), CASS Foundation (B.G.F.), Commonwealth of Australia Department of Health and Aging (B.G.F.), Ian Potter Foundation (B.G.F.), International Human Frontiers Science Program Organisation (B.G.F.), Leiden University (F.J.V., M.K.R.), NASA Astrobiology Institute (S.B.H.), National Science Foundation (S.B.H.) and University of Melbourne (B.G.F.). We thank the relevant wildlife departments for granting the scientific permits for field collection of required specimens.

Author Information The sequences of the cDNA clones have been deposited in GenBank (accession numbers DQ139877–DQ139931 and DQ184481), as have the nuclear gene sequences (DQ119594–DQ119641). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.G.F. (bgf@unimelb.edu.au).

GABA regulates synaptic integration of newly generated neurons in the adult brain

Shaoyu Ge^{1*}, Eyleen L. K. Goh^{1*}, Kurt A. Sailor¹, Yasuji Kitabatake¹, Guo-li Ming¹ & Hongjun Song¹

Adult neurogenesis, the birth and integration of new neurons from adult neural stem cells, is a striking form of structural plasticity and highlights the regenerative capacity of the adult mammalian brain^{1–8}. Accumulating evidence suggests that neuronal activity regulates adult neurogenesis and that new neurons contribute to specific brain functions^{1–8}. The mechanism that regulates the integration of newly generated neurons into the pre-existing functional circuitry in the adult brain is unknown. Here we show that newborn granule cells in the dentate gyrus of the adult hippocampus are tonically activated by ambient GABA (γ -aminobutyric acid) before being sequentially innervated by GABA- and glutamate-mediated synaptic inputs. GABA, the major inhibitory neurotransmitter in the adult brain, initially exerts an excitatory action on newborn neurons owing to their high cytoplasmic chloride ion content^{9–12}. Conversion of GABA-induced depolarization (excitation) into hyperpolarization (inhibition) in newborn neurons leads to marked defects in their synapse formation and dendritic development *in vivo*. Our study identifies an essential role for GABA in the synaptic integration of newly generated neurons in the adult brain, and suggests an unexpected mechanism for activity-dependent regulation of adult neurogenesis, in which newborn neurons may sense neuronal network activity through tonic and phasic GABA activation.

Using a retroviral strategy to express green fluorescent protein (GFP) specifically in proliferating cells and their progeny^{6,7}, we examined the synaptic integration of newly generated granule cells (DGCs) in the dentate gyrus of adult mice (Fig. 1). Retroviral labelling provides adequate time resolution for birth-dating and does not seem to affect the neuronal development of labelled cells (Fig. 1a, Supplementary Fig. 1, Supplementary Videos 1, 2 and Supplementary Table 1). To monitor the integration of new neurons in the adult brain, we used whole-cell patch clamping and recorded from GFP⁺ DGCs in slices prepared from virus-infected animals (see Methods). Three days post viral injection (3 dpi), none of the GFP⁺ cells recorded under voltage-clamp ($V_m = -65$ mV) showed any spontaneous synaptic currents (SSCs) or any detectable evoked postsynaptic currents (PSCs) when the perforant pathway was stimulated ($n = 15$; Fig. 1b–d). However, bath application of bicuculline (100 μ M), a specific GABA_A receptor antagonist^{13,14}, revealed the presence of a tonic current in all GFP⁺ DGCs recorded from 3 dpi and onwards ($n = 48$; Fig. 1b). SR95531 (100 μ M), another GABA_A receptor antagonist^{13,14}, also abolished the tonic current (Supplementary Fig. 2a). In contrast, NO-711 (2.5 μ M), a specific GABA transporter inhibitor^{13,14}, significantly enhanced the tonic current (Supplementary Fig. 2b). Notably, stimulation of local interneurons, such as basket cells¹⁵, also enhanced the tonic currents in newborn DGCs (Supplementary Fig. 2c). Thus, newborn DGCs in the adult brain are tonically activated by ambient GABA before any detectable phasic/synaptic activation. Bicuculline (10 μ M)-sensitive GABA-mediated

PSCs (Fig. 1c) and CNQX (50 μ M)-sensitive glutamate-mediated PSCs (Fig. 1d) were first detected in some GFP⁺ DGCs at 7 dpi and 14 dpi, respectively. These results show that newborn neurons in the adult brain, as in neonates, follow a stereotypical integration process—first receiving tonic GABA activation, then GABA-mediated synaptic inputs and finally glutamate-mediated synaptic inputs^{9,10,16–20}.

To determine the nature of GABA activation, we made perforated whole-cell patch-clamp recordings using gramicidin (25 μ g ml⁻¹) to allow reliable recording of GABA-induced currents²¹. We found that the reversal potential for GABA-induced currents (E_{GABA}) in GFP⁺ DGCs gradually decreased during maturation (Fig. 2a, Supplementary Fig. 3a), indicating a higher concentration of intracellular chloride [Cl^-]_i in younger neurons (Supplementary Fig. 4). However, the resting membrane potential (V_{rest}) decreased only slightly over time (Fig. 2a and Supplementary Fig. 3b). Notably, V_{rest} was significantly more negative than E_{GABA} during the first two weeks after viral injection (Fig. 2a). Thus, GABA initially depolarizes newborn DGCs in the adult brain.

The polarity of GABA action is largely determined by neuronal [Cl^-]_i (refs 9–12). Sequential expression of the Na⁺-K⁺-2Cl⁻ transporter NKCC1 (a Cl⁻ importer) and the K⁺-coupled Cl⁻ transporter KCC2 (a Cl⁻ exporter) is believed to underlie the conversion from depolarization to hyperpolarization by GABA during neuronal maturation in the fetal brain^{9–12}. We found that newborn DGCs (doublecortin-positive) in the adult brain express high levels of NKCC1 and little KCC2 (Fig. 2b and Supplementary Fig. 5b, c). We constructed several retroviruses expressing specific short hairpin RNAs (shRNA) against different regions of mouse *Nkcc1* (also known as *Slc12a2*) (ref. 22). We found that two different *Nkcc1*-shRNAs, but not the control shRNA, almost completely knocked down the expression of NKCC1 as shown by western blot analysis of HEK293 cells (Supplementary Fig. 5a) and immunostaining of newborn DGCs (Fig. 2b and Supplementary Fig. 5b). None of these shRNAs affected KCC2 expression in the infected cells *in vivo* (Fig. 2b and Supplementary Fig. 5c). GFP⁺ DGCs expressing *Nkcc1*-shRNA, but not the control shRNA, had significantly lower [Cl^-]_i (Supplementary Fig. 4). In addition, E_{GABA} was more negative than V_{rest} in the *Nkcc1*-shRNA-expressing DGCs throughout their development (Fig. 2c). Under perforated patch recording using gramicidin, tonic GABA activation led to hyperpolarization of *Nkcc1*-shRNA-expressing DGCs at 7 dpi, in contrast to depolarization of the control newborn DGCs (Fig. 2d). The amplitude of the tonic GABA currents under whole-cell recording, however, was similar (Fig. 2d), suggesting that the expression levels of functional GABA_A receptors responsible for the tonic activation was not significantly affected.

We next examined the synaptic integration of new DGCs in the absence of GABA-induced depolarization *in vivo*. GABA-mediated

¹Institute for Cell Engineering, Departments of Neurology and Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

*These authors contributed equally to this work.

synaptic transmission was examined in the presence of kynurenic acid (5 mM), which blocks ionotropic glutamate-mediated currents (Fig. 3a–c). We could not detect any PSCs in *Nkcc1*-shRNA-expressing DGCs at 7 dpi (Fig. 3a, b). The mean amplitude of the recorded PSCs at 14 and 28 dpi was only about 12% and 65% of those observed in control GFP⁺ DGCs, respectively (Fig. 3b). In addition, the frequency of SSCs recorded at 28 dpi, but not the mean amplitude, was also significantly reduced (Fig. 3c), further indicating defects in the formation of GABA-mediated synapses by *Nkcc1*-shRNA-expressing cells.

We then examined glutamate-mediated synaptic transmission in the presence of bicuculline (10 μ M) to block ionotropic GABA-mediated currents (Fig. 3d–f). We could not detect any PSCs or SSCs in *Nkcc1*-shRNA-expressing DGCs at 14 dpi, and the percentage of cells recorded with PSCs was greatly reduced at 28 dpi (Fig. 3d, e). Moreover, the mean peak amplitude of PSCs and frequency of SSCs at 28 dpi were only about 42% and 7% of those from control GFP⁺ DGCs, respectively (Fig. 3e, f). The mean amplitude of SSCs, however, was not significantly different (Fig. 3f), suggesting that there were no general defects in receptor expression at the synapses.

We also examined the synaptic integration of newborn DGCs in *Nkcc1* germline knockout (*Nkcc1*^{−/−}) mice²³. Despite the caveats of defects and potential compensation during embryonic development¹¹, we found similar defects in the formation of GABA- and glutamate-mediated synapses by newborn DGCs in adult *Nkcc1*^{−/−} mice (Supplementary Fig. 6). Taken together, these results suggest

that GABA-induced depolarization is essential for the establishment of functional GABA- and glutamate-mediated synapses for newly generated DGCs in the adult brain.

To directly examine the functional role of GABA-induced depolarization in the structural plasticity of newborn neurons, we used confocal microscopy to reconstruct the dendritic arborization of GFP⁺ DGCs at 14 dpi (Fig. 4a), when active synaptogenesis occurs for GABA- and glutamate-mediated inputs. Consistent with results from the electrophysiological studies, we found that *Nkcc1*-shRNA-expressing DGCs had marked defects in dendritic arborization (Fig. 4b, c). The total dendritic length and branch number of these new neurons, as well as their dendritic complexity, were significantly reduced. In addition, we found that injection of the GABA_A receptor agonist pentobarbital²⁰ seems to promote dendrite growth of newborn DGCs *in vivo* (Supplementary Fig. 7). Thus, GABA-induced depolarization/excitation regulates the dendritic development of newborn neurons in the adult brain.

Combining electrophysiology with retrovirus-mediated birth-dating and labelling, we have delineated the sequential steps involved in the integration of newly generated neurons into the pre-existing functional circuitry in the adult brain: from tonic GABA activation to GABA-mediated synaptic innervation and finally glutamate-mediated synaptic innervation (Figs 1, 3). GABA exerts a depolarizing action during the initial development of new DGCs due to their high [Cl[−]]_i from the expression of NKCC1 (Fig. 2). Using a retrovirus-mediated 'single-cell genetic' approach, we have shown that converting GABA-induced depolarization into hyperpolarization led

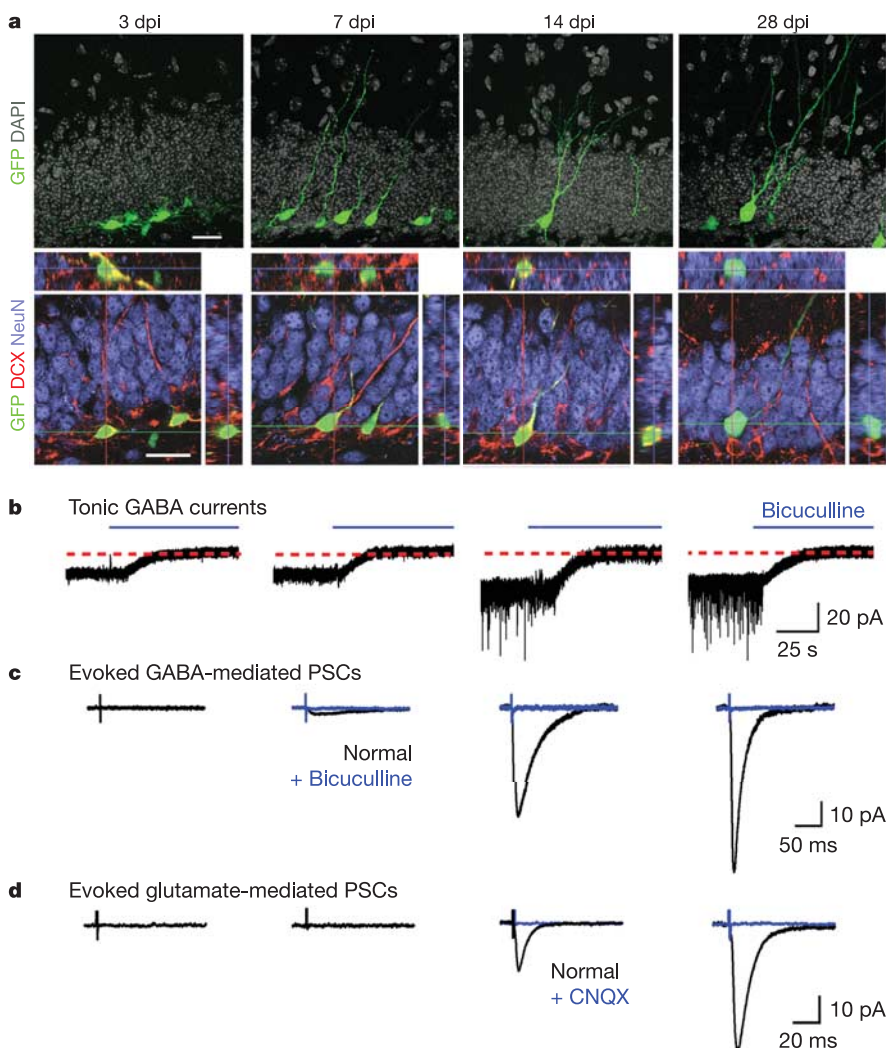


Figure 1 | Development of newborn DGCs in the adult mice. **a**, Confocal images of new DGCs (GFP⁺, green) at different stages. Shown are projections (top) and confocal images of immunostaining (bottom) for doublecortin (DCX, red) and neuron-specific nuclear protein (NeuN, blue) with orthogonal views to confirm the co-localization of GFP and DCX or NeuN. Scale bars, 20 μ m. **b–d**, Synaptic integration of newborn DGCs. Shown are sample recording traces from GFP⁺ DGCs under whole-cell voltage-clamp ($V_m = -65$ mV). Tonic currents shown (**b**) are continuous recordings before and after adding bicuculline (100 μ M, blue). Evoked PSCs shown are averaged responses from 5 consecutive stimuli before (black) and after (blue) adding bicuculline (10 μ M, **c**) or CNQX (50 μ M, **d**), respectively.

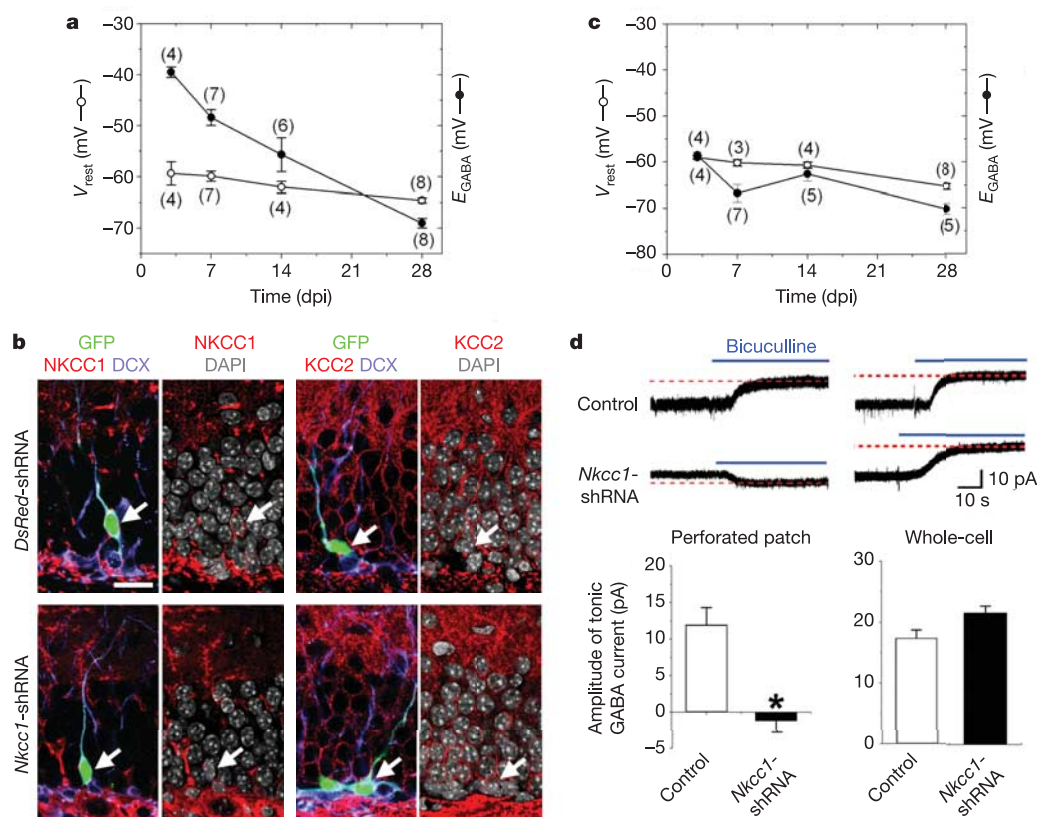


Figure 2 | Nature of GABA-induced activation in newborn DGCs in the adult brain. **a**, Resting membrane potentials (V_{rest}) and GABA-reversal potentials (E_{GABA}) of GFP⁺ DGCs. Data show mean $\pm$ s.e.m. Numbers associated with symbols refer to the number of cells examined. **b**, Retrovirus-mediated co-expression of GFP and shRNAs specific for *Nkcc1*, but not a control shRNA (*DsRed*-shRNA), reduced NKCC1 expression but had no effects on KCC2 expression in newborn DGCs (7 dpi). Shown are confocal images of

GFP (green) and immunostaining for NKCC1 or KCC2 (red), DCX (blue) and DAPI (grey). Arrows point to GFP⁺ DGCs. Scale bar, 20 μ m. **c**, V_{rest} and E_{GABA} in *Nkcc1*-shRNA-expressing newborn DGCs (similar to **a**). **d**, Tonic GABA currents in newborn DGCs (7 dpi) recorded under perforated patch (with gramicidin) or break-in whole-cell recording ($V_m = -65$ mV). Blue lines indicate the addition of bicuculline (100 μ M). Values in bar graphs represent mean $\pm$ s.e.m. ($n = 6$, * $P < 0.01$, ANOVA).

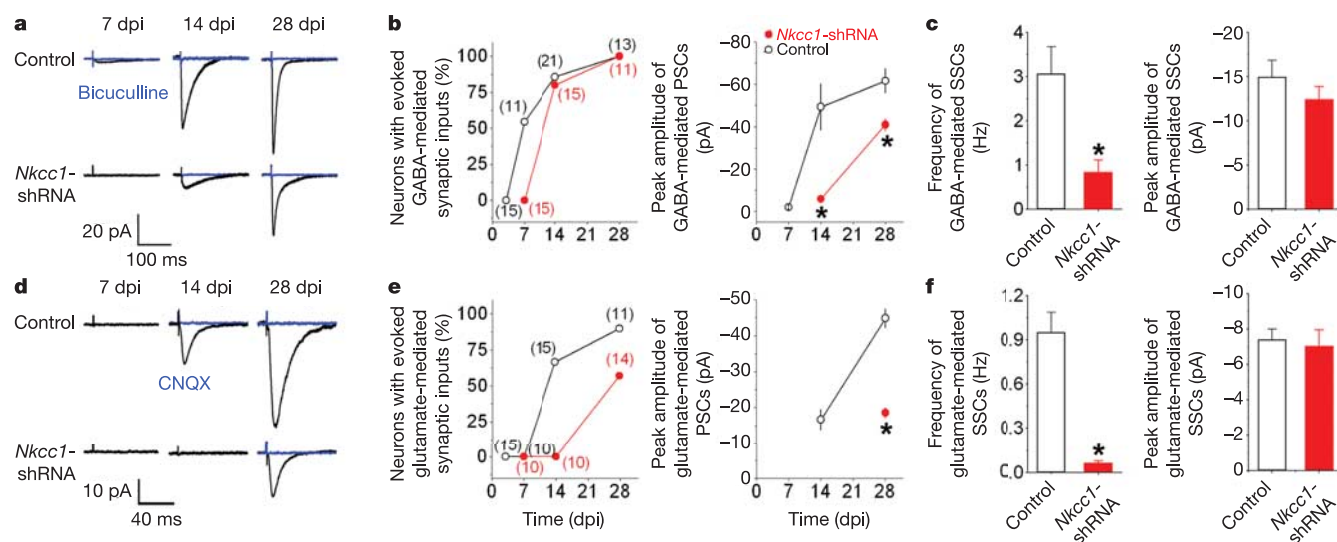


Figure 3 | Synaptic integration of newborn DGCs in the adult brain. **a–c**, Formation of GABA-mediated synaptic inputs by GFP⁺ DGCs. Shown in **a** are sample traces of evoked PSCs recorded under whole-cell voltage-clamp ($V_m = -65$ mV, 5 mM kynurenic acid in the bath) before and after the addition of bicuculline (10 μ M). **b**, **c**, The percentages of GFP⁺ DGCs with detectable GABA-mediated PSCs, the mean peak amplitude of GABA-mediated PSCs (**b**), and mean frequency and peak amplitude of

GABA-mediated SSCs recorded at 28 dpi (**c**). Numbers associated with symbols refer to the number of cells examined. Values represent mean $\pm$ s.e.m. (* $P < 0.01$, ANOVA). **d–f**, Formation of glutamate-mediated synaptic inputs by GFP⁺ DGCs. Same as **a–c**, except that the recordings were carried out in the presence of bicuculline (10 μ M). Blue lines in **d** indicate the addition of CNQX (50 μ M).

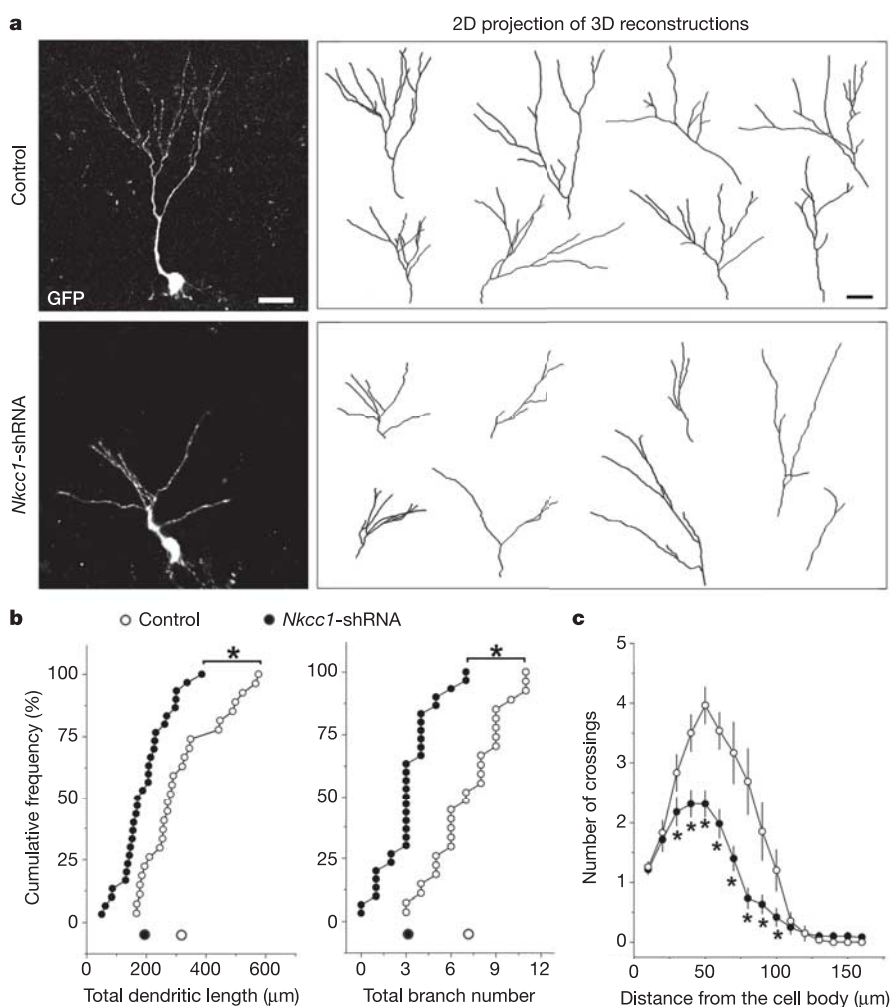


Figure 4 | Dendritic development of newborn DGCs in the adult brain. **a**, Confocal three-dimensional (3D) reconstruction of dendrites of control or *Nkcc1*-shRNA-expressing DGCs (14 dpi). Scale bar, 20 μm. **b**, Quantification of total dendritic length and branch number of newborn DGCs. Each symbol represents data from a single control (empty) or *Nkcc1*-shRNA-expressing (filled) DGC at 14 dpi. Symbols on the x axis represent mean values. (* $P < 0.01$, Kolmogorov–Smirnov test). **c**, Sholl analysis of dendritic complexity of GFP⁺ DGCs (14 dpi). Values represent mean $\pm$ s.e.m. ($n = 27$; * $P < 0.05$, Student's *t*-test).

to significant defects in the formation of GABA- and glutamate-mediated synapses as well as in the dendritic development of newly generated neurons in the adult brain (Figs 3, 4). In the adult brain, ambient GABA is known to regulate the excitability of certain mature neurons, notably in the cerebellum and dentate gyrus^{13–15,24,25}. Here we have shown that tonic GABA activation depolarizes newborn DGCs (Fig. 2 and Supplementary Fig. 3c), and more importantly, it constitutes the bulk of GABA-induced activation during the initial integration process when the phasic GABA activation either does not exist or is weaker than the tonic activation (Figs 2d and 3b, c). The mechanism by which tonic GABA activation regulates neuronal development and synaptic integration of new DGCs in the adult brain remains to be determined. Both voltage-dependent⁸ and -independent Ca^{2+} -permeable channels²⁶ could be involved. Newborn DGCs in the adult brain express high levels of low-voltage activated T-type Ca^{2+} channels that are activated below -57 mV (ref. 8). Thus, tonic depolarization by GABA may lead to an activation of these Ca^{2+} channels and subsequent Ca^{2+} influx. Tonic activation may also provide an initial depolarization that allows a small phasic GABA activation to reach the threshold of these Ca^{2+} channels.

Activity-dependent anatomical reorganization is widely regarded as a fundamental mechanism of developmental and adult neural plasticity^{27,28}. Within the dentate gyrus, principal neurons and interneurons form extensive recurrent connections. The levels of ambient GABA, regulated by interneuron activities (Supplementary Fig. 2c), may serve as a general indicator of dynamic neuronal network activity. Our study thus suggests an unexpected mechanism for activity-dependent regulation of adult neurogenesis, in which

newborn neurons, before receiving any synaptic innervations, may sense neuronal network activities through local ambient GABA levels. Many physiological and pathological stimulations, including neurosteroids and epilepsy, affect GABA signalling^{15,25} and might therefore influence the integration of new neurons in the adult brain. Our study might also have important implications for the use of stem cells in neuronal cell-replacement therapy for degenerative neurological diseases.

METHODS

See Supplementary Information for detailed methods.

Construction, production and stereotaxic injection of engineered retroviruses. Engineered self-inactivating murine retroviruses were used to label and genetically manipulate proliferating cells and their progeny^{6,7}. GFP and shRNA were co-expressed under the control of the EF1 α and human U6 promoters²², respectively. The following short hairpin sequences were used: ACACACTTGTCTCTGGGATT (*Nkcc1*-shRNA1); GGACAATATCTACCCAGCT (*Nkcc1*-shRNA2); AGTTCCAGTACGGCTCCAA (DsRed-shRNA). The specificity and efficiency of the shRNAs were validated, and high titres of engineered retroviruses (1×10^9 units ml^{-1}) were produced as previously described⁶.

Adult (7–8-week-old) female C57BL/6 mice (Charles River) and *Nkcc1*^{−/−} mice²³ housed under standard conditions were anaesthetized and retroviruses were stereotaxically injected at four sites ($0.5 \mu\text{l}$ per site at $0.25 \mu\text{l min}^{-1}$) with the following coordinates (from bregma in mm), as previously described⁶: anteroposterior, -2 ; lateral, ± 1.6 ; ventral, 2.5 ; and anteroposterior, -3 ; lateral, ± 2.6 ; ventral, 3.2 . A total of 530 animals were used and all animal procedures were treated in accordance with institutional guidelines.

Immunostaining, confocal imaging and analysis. Coronal brain sections ($40 \mu\text{m}$ thick) were prepared and processed for immunostaining using the following antibodies, as previously described⁶: goat anti-DCX (Santa Cruz; 1:500), mouse anti-NeuN (Chemicon; 1:200), mouse anti-NKCC1 (T4,

Developmental Studies Hybridoma Bank; 1:200), rabbit anti-KCC2 (Upstate; 1:200) and rabbit anti-Ki67 (Novocastra; 1:500). Sections were also stained for 4',6-diamidino-2-phenylindole (DAPI; 1:5,000). Images were acquired on a Zeiss LSM 510 META multiphoton confocal system using a multi-track configuration. For dendritic analysis, three-dimensional reconstructions of the dendritic processes of each GFP⁺ neuron were made from Z-series stacks of confocal images. The projection images were semi-automatically traced with NIH ImageJ using the NeuronJ plugin. Total dendritic length and branch number of each individual GFP⁺ neuron in the granule cell layer were analysed. Statistical significance ($P < 0.01$) was assessed using the Kolmogorov–Smirnov test. The Sholl analysis for dendritic complexity was carried out by counting the number of dendrites that cross a series of concentric circles at 5- μ m intervals from the soma. Statistical significance ($P < 0.05$) was assessed using Student's *t*-test.

Electrophysiology. Mice housed under standard conditions were processed for slice preparation and electrophysiology as previously described⁶. Electrophysiological recordings were obtained at 32–34°C. GFP⁺ DGCs were identified by their green fluorescence, location within the subgranule or granule cell layer, neuronal morphology and capacity to generate Na⁺ spikes (7 dpi and onwards). We monitored V_{rest} based on the reversal potential of the K⁺ current through cell-attached patches (Supplementary Fig. 3b), to avoid underestimation of V_{rest} due to a shunt through the seal contact between the pipette and the membrane in perforated and whole-cell recording^{18,29,30}. Microelectrodes (4–6 M Ω) were filled with the following solution: 120 mM potassium gluconate, 15 mM KCl, 4 mM MgCl₂, 0.1 mM EGTA, 10.0 mM HEPES, 4 mM MgATP, 0.3 mM Na₃GTP, 7 mM phosphocreatine (pH 7.4, 300 mOsm). For characterizing tonic GABA currents, potassium salt was substituted with CsCl in the intracellular solution and tetrodotoxin (0.5 μ M) was added to the recording solution¹⁴. Additional drugs were used at following final concentrations: bicuculline (100 μ M, Sigma), SR95531 (100 μ M, Tocris), NO-711 (2.5 μ M, Sigma).

Data were collected using an Axon 200B amplifier and acquired with a DigiData 1322A (Axon Instruments) at 10 kHz. Series and input resistances were monitored, and only those with changes less than 20% during experiments were analysed. The series resistance ranged from 10–30 M Ω and was uncompensated. For perforated patch recordings, the gramicidin stock (10 mg ml⁻¹ in DMSO) was diluted in the pipette solution (135 mM CsCl, 4 mM MgCl₂, 0.1 mM EGTA, 10 mM HEPES, pH 7.4, 300 mOsm) to a final concentration of 25 μ g ml⁻¹ just before experiments. Perforated patch recordings with a series resistance of <80 M Ω and without significant changes (>25%) during recordings were used for data analysis.

For measurement of E_{GABA} , focal pressure ejection of 10 μ M GABA through a puffer pipette controlled by a Picospitzer (5-ms puff at 3–5 psi) was used to activate GABA_A receptors on the GFP⁺ DGCs with gramicidin perforated patch under voltage-clamp at different holding potentials (Supplementary Fig. 3a). The peak amplitude and holding potential were plotted and E_{GABA} was determined for each cell. Intracellular chloride concentrations were calculated using the following equation: $[Cl^-]_i = [Cl^-]_o e^{(E_{GABA} - F/RT)}$, where the extracellular Cl⁻ concentration $[Cl^-]_o$ is 134.1 mM (Supplementary Fig. 4).

A bipolar electrode (World Precision Instruments) was used to stimulate (100- μ s duration) the perforant pathway input to the dentate gyrus. The stimulus intensity (~30 μ A) was maintained for all experiments. To examine the evoked synaptic transmission, a train of 20 stimuli was delivered at 0.1 Hz. To confirm a lack of evoked synaptic transmission, the stimulation intensity was then increased to 200 μ A.

Received 1 September; accepted 8 November 2005.

Published online 11 December 2005.

- Kempermann, G. & Gage, F. H. New nerve cells for the adult brain. *Sci. Am.* **280**, 48–53 (1999).
- Fuchs, E. & Gould, E. Mini-review: *in vivo* neurogenesis in the adult brain: regulation and functional implications. *Eur. J. Neurosci.* **12**, 2211–2214 (2000).
- Temple, S. & Alvarez-Buylla, A. Stem cells in the adult mammalian central nervous system. *Curr. Opin. Neurobiol.* **9**, 135–141 (1999).
- Doetsch, F. & Hen, R. Young and excitable: the function of new neurons in the adult mammalian brain. *Curr. Opin. Neurobiol.* **15**, 121–128 (2005).
- Ming, G.-I. & Song, H. Adult neurogenesis in the mammalian central nervous system. *Annu. Rev. Neurosci.* **28**, 223–250 (2005).
- van Praag, H. *et al.* Functional neurogenesis in the adult hippocampus. *Nature* **415**, 1030–1034 (2002).
- Carleton, A., Petreanu, L. T., Lansford, R., Alvarez-Buylla, A. & Lledo, P. M. Becoming a new neuron in the adult olfactory bulb. *Nature Neurosci.* **6**, 507–518 (2003).
- Schmidt-Hieber, C., Jonas, P. & Bischofberger, J. Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature* **429**, 184–187 (2004).

- Ben-Ari, Y. Excitatory actions of GABA during development: the nature of the nurture. *Nature Rev. Neurosci.* **3**, 728–739 (2002).
- Owens, D. F. & Kriegstein, A. R. Is there more to GABA than synaptic inhibition? *Nature Rev. Neurosci.* **3**, 715–727 (2002).
- Delpire, E. Cation-chloride cotransporters in neuronal communication. *News Physiol. Sci.* **15**, 309–312 (2000).
- Payne, J. A., Rivera, C., Voipio, J. & Kaila, K. Cation-chloride co-transporters in neuronal communication, development and trauma. *Trends Neurosci.* **26**, 199–206 (2003).
- Overstreet, L. S. & Westbrook, G. L. Paradoxical reduction of synaptic inhibition by vigabatrin. *J. Neurophysiol.* **86**, 596–603 (2001).
- Nusser, Z. & Mody, I. Selective modulation of tonic and phasic inhibitions in dentate gyrus granule cells. *J. Neurophysiol.* **87**, 2624–2628 (2002).
- Farrant, M. & Nusser, Z. Variations on an inhibitory theme: phasic and tonic activation of GABA_A receptors. *Nature Rev. Neurosci.* **6**, 215–229 (2005).
- Wang, L. P., Kempermann, G. & Kettenmann, H. A subpopulation of precursor cells in the mouse dentate gyrus receives synaptic GABAergic input. *Mol. Cell. Neurosci.* **29**, 181–189 (2005).
- Wadiche, L. O., Bromberg, D. A., Bensen, A. L. & Westbrook, G. L. GABAergic signalling to newborn neurons in dentate gyrus. *J. Neurophysiol.* **94**, 4528–4532 (2005).
- Wang, D. D., Krueger, D. D. & Bordey, A. GABA depolarizes neuronal progenitors of the postnatal subventricular zone via GABA_A receptor activation. *J. Physiol. (Lond.)* **550**, 785–800 (2003).
- Liu, X., Wang, Q., Haydar, T. F. & Bordey, A. Nonsynaptic GABA signalling in postnatal subventricular zone controls proliferation of GFAP-expressing progenitors. *Nature Neurosci.* **8**, 1179–1187 (2005).
- Tozuka, Y., Fukuda, S., Namba, T., Seki, T. & Hisatsune, T. GABAergic excitation promotes neuronal differentiation in adult hippocampal progenitor cells. *Neuron* **47**, 803–815 (2005).
- Owens, D. F., Boyce, L. H., Davis, M. B. & Kriegstein, A. R. Excitatory GABA responses in embryonic and neonatal cortical slices demonstrated by gramicidin perforated-patch recordings and calcium imaging. *J. Neurosci.* **16**, 6414–6423 (1996).
- Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
- Delpire, E., Lu, J., England, R., Dull, C. & Thorne, T. Deafness and imbalance associated with inactivation of the secretory Na-K-2Cl co-transporter. *Nature Genet.* **22**, 192–195 (1999).
- Chadderton, P., Margrie, T. W. & Häusser, M. Integration of quanta in cerebellar granule cells during sensory processing. *Nature* **428**, 856–860 (2004).
- Semyanov, A., Walker, M. C., Kullmann, D. M. & Silver, R. A. Tonically active GABA_A receptors: modulating gain and maintaining the tone. *Trends Neurosci.* **27**, 262–269 (2004).
- Chavas, J., Forero, M. E., Collin, T., Llano, I. & Marty, A. Osmotic tension as a possible link between GABA_A receptor activation and intracellular calcium elevation. *Neuron* **44**, 701–713 (2004).
- Cline, H. T. Dendritic arbor development and synaptogenesis. *Curr. Opin. Neurobiol.* **11**, 118–126 (2001).
- Wong, R. O. & Ghosh, A. Activity-dependent regulation of dendritic growth and patterning. *Nature Rev. Neurosci.* **3**, 803–812 (2002).
- Tyzio, R. *et al.* Membrane potential of CA3 hippocampal pyramidal cells during postnatal development. *J. Neurophysiol.* **90**, 2964–2972 (2003).
- Verheugen, J. A., Fricker, D. & Miles, R. Noninvasive measurements of the membrane potential and GABAergic action in hippocampal interneurons. *J. Neurosci.* **19**, 2546–2555 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank C. F. Stevens, F. H. Gage, R. Huganir, K.-W. Yau and J. Bischofberger for comments and suggestions, L.-h. Liu for technical support, E. Delpire for *Nkcc1* knockout mice and mouse *Nkcc1* cDNA, and N. Gaiano, D. Sun, D. K. Ma and D. Pradhan for reagents and help. This work was supported by the National Institute of Health (H.S.), Klingenstein Fellowship Awards in the Neurosciences (G.-I.M. and H.S.), the Whitehall Foundation (G.-I.M.) and The Robert Packard Center for ALS Research at Johns Hopkins (H.S.).

Author Contributions S.G. did virus injection and electrophysiology, E.L.K.G. engineered retroviral constructs and did characterization, K.A.S. did immunohistochemistry and confocal imaging analysis, and Y.K. helped with molecular biology. G.-I.M. and H.S. are senior authors and were responsible for project planning. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at ng.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.S. (shongju1@bs.jhmi.edu) or G.-I.M. (gming1@bs.jhmi.edu).

LETTERS

Lgl, Pins and aPKC regulate neuroblast self-renewal versus differentiation

Cheng-Yu Lee¹, Kristin J. Robinson¹ & Chris Q. Doe¹

How a cell chooses to proliferate or to differentiate is an important issue in stem cell and cancer biology. *Drosophila* neuroblasts undergo self-renewal with every cell division, producing another neuroblast and a differentiating daughter cell, but the mechanisms controlling the self-renewal/differentiation decision are poorly understood. Here we tested whether cell polarity genes, known to regulate embryonic neuroblast asymmetric cell division¹, also regulate neuroblast self-renewal. Clonal analysis in larval brains showed that *pins* mutant neuroblasts rapidly fail to self-renew, whereas *lethal giant larvae* (*lgl*) mutant neuroblasts generate multiple neuroblasts. Notably, *lgl pins* double mutant neuroblasts all divide symmetrically to self-renew, filling the brain with neuroblasts at the expense of neurons. The *lgl pins* neuroblasts show ectopic cortical localization of atypical protein kinase C (aPKC), and a decrease in aPKC expression reduces neuroblast numbers, suggesting that aPKC promotes neuroblast self-renewal. In support of this hypothesis, neuroblast-specific overexpression of membrane-targeted aPKC, but not a kinase-dead version, induces ectopic neuroblast self-renewal. We conclude that cortical aPKC kinase activity is a potent inducer of neuroblast self-renewal.

Drosophila neuroblasts are an excellent model system in which to investigate the molecular control of self-renewal versus differentiation. Larval neuroblasts repeatedly divide asymmetrically to

self-renew a neuroblast and to produce a smaller daughter cell, called a ganglion mother cell (GMC), that typically makes two postmitotic neurons; this process enables a single neuroblast to generate many hundreds of neurons (Fig. 1a). We define self-renewal as the capacity of a neuroblast to maintain all attributes of its cell type (molecular markers and proliferation potential). In this regard, a neuroblast is very similar to a germline stem cell: both maintain their stem cell identity while generating differentiating progeny². About 100 neuroblasts per brain lobe are formed during embryogenesis, where they proliferate briefly before entering quiescence³. Brain neuroblasts re-enter the cell cycle between 10 and 72 h after larval hatching (ALH)⁴, and then a stable population of ~100 mitotic, self-renewing neuroblasts is maintained (Fig. 1b). We are using this invariant neuroblast number to screen for mutants altering self-renewal versus differentiation: mutants in which a neuroblast makes two neuroblast progeny (ectopic self-renewal) will have >100 neuroblasts, whereas mutants in which a neuroblast makes two GMC progeny (failure in self-renewal) will have <100 neuroblasts. Here we have used this assay to test known cell polarity mutants for a role in neuroblast self-renewal; the results of larger genetic screens will be described elsewhere.

We assayed two classes of cell polarity regulators for an effect on larval neuroblast self-renewal. We first examined *lgl* and *discs large* (*dlg*) zygotic mutants, because these mutants form brain tumours⁵

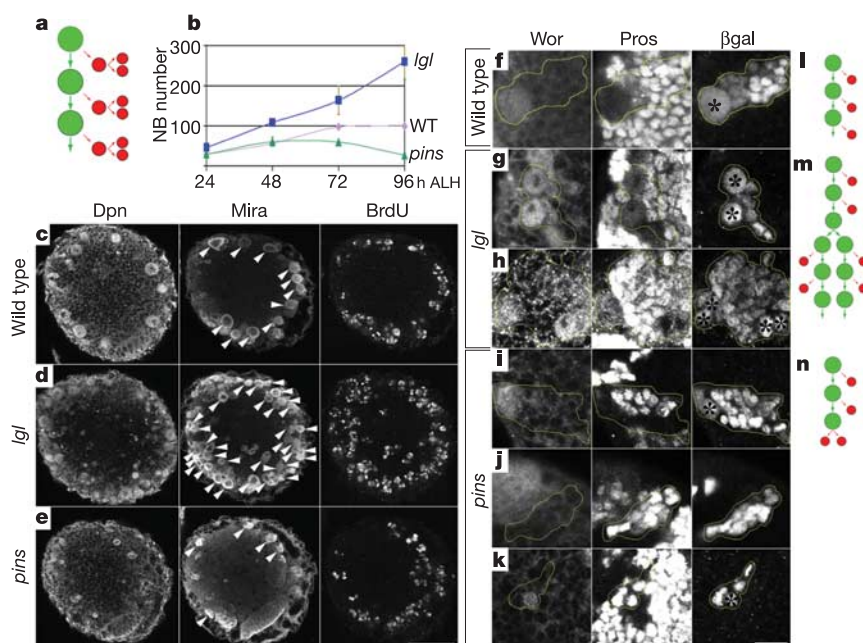


Figure 1 | *lgl* and *pins* regulate larval neuroblast self-renewal.

a, Wild-type neuroblast lineage. Green indicates neuroblast self-renewal, red indicates GMC differentiation. **b**, Quantification of neuroblast numbers from 24 to 96 h after larval hatching (ALH) ($n = 50$ per time point per genotype). **c–e**, Single focal plane through wild-type (**c**), *lgl*³³⁴ (**d**) and *pins*⁶² (**e**) brains at 96 h ALH stained with the indicated markers (neuroblasts are indicated in one panel with arrowheads). **f–k**, Clonal analysis of single neuroblast lineages. Single neuroblast clones marked with β -galactosidase are circled; asterisks indicate neuroblasts. **f**, In wild type, neuroblast clones always contain a single *Worniu*⁺ neuroblast and several *Prospero*⁺ progeny (not all are shown). **g, h**, In *lgl*³³⁴ mutants, neuroblast clones typically show two neuroblasts (**g**) or as many as six neuroblasts (**h**; only four are in the focal plane). **i–k**, In *pins*⁶² mutants, neuroblast clones either have one (**i**) or zero (**j**) neuroblasts; in one clone, we observed a cell with both neuroblast and GMC markers (**k**). **l–m**, Interpretation of wild type and mutant lineages. Scale bars, 50 μ m (**c–e**); 10 μ m (**f–k**).

¹Institutes of Neuroscience and Molecular Biology, Howard Hughes Medical Institute, University of Oregon 1254, Eugene, Oregon 97403, USA.

and promote basal protein targeting in embryonic and larval neuroblasts^{6–8}. *Lgl* and *Dlg* have several protein interaction motifs and are localized around the neuroblast cortex^{6–10}. In addition, we examined *pins* and *Gai* zygotic mutants; these genes regulate cell polarity in embryonic neuroblasts¹, but have not been well characterized in larval neuroblasts. *Pins* and *Gai* are colocalized with Inscuteable and the evolutionarily conserved Bazooka–Par6–aPKC proteins at the apical cortex of mitotic neuroblasts, and all of these proteins are partitioned into the neuroblast during cytokinesis¹.

In wild-type larvae, a population of ~100 neuroblasts could be identified by the markers Worniu, Deadpan and Miranda, and by labelling with a pulse of 5-bromodeoxyuridine (BrdU); by contrast, the thousands of differentiating GMCs and neurons rapidly down-regulated neuroblast markers and expressed nuclear Prospero and/or Elav (Fig. 1b, c, and Supplementary Fig. 1a and Movie 1). We observed a clear increase in neuroblast number in *lgl* and *dlg* mutants; there were supernumerary neuroblasts at all stages examined, and all extra neuroblasts expressed Deadpan and Miranda and were proliferative on the basis of their ability to incorporate BrdU (Fig. 1b, d, Supplementary Fig. 1 and Movie 2, and data not shown). *Gai* zygotic mutants had a complex phenotype that will be described elsewhere; however, *pins* zygotic mutants showed a marked decrease in neuroblast number (Fig. 1b, e, and Supplementary Fig. 1 and Movie 3). Notably, this phenotype was not due to a subset of neuroblasts remaining quiescent, because neuroblast numbers peaked and then declined over time (Fig. 1b), and it was not due to neuroblast cell death (see below). The relatively late onset of the *pins* phenotype was probably due to the gradual depletion of maternal *pins* gene product in these larvae (C.-Y.L. and C.Q.D., unpublished data).

To determine whether the *pins* and *lgl* larval brain phenotypes were due to defects in neuroblast self-renewal, we induced positively marked genetic clones¹¹ in single neuroblasts to trace their progeny (Fig. 1f–k). We adjusted clone induction parameters to ensure that each clone was derived from a single neuroblast (1.2 clones per lobe; $n = 20$). In wild-type brains, neuroblast clones always contained a single Worniu⁺ Miranda⁺ nuclear-Prospero[–] neuroblast and numerous smaller Worniu[–] Miranda[–] nuclear-Prospero⁺ progeny (Fig. 1f; $n = 35$ clones), confirming that wild-type neuroblasts always divide to self-renew and to generate a smaller differentiating GMC (Fig. 1i). By contrast, *lgl* mutant brains had an average of 2.3 neuroblasts per clone, with up to six neuroblasts per clone (Fig. 1g, h; $n = 22$ clones), showing that *lgl* mutant neuroblasts can divide symmetrically to yield two neuroblasts (Fig. 1m). The opposite phenotype was seen in *pins* mutant brains: 72.8% of the clones had no neuroblast and the remainder had a single neuroblast (Fig. 1i–k; $n = 34$ clones). The neuroblasts did not die in the *pins* mutants as the cell death marker caspase-3 was not upregulated (Supplementary Fig. 2), as neuroblast-specific expression of the p35 cell death inhibitor did not rescue the missing neuroblasts (data not shown), and as we observed one clone in which the largest cell coexpressed neuroblast and GMC markers, consistent with an intermediate stage in neuroblast-to-GMC differentiation (Fig. 1k). We conclude that wild-type neuroblasts exclusively generate neuroblast/GMC siblings; *lgl* mutant neuroblasts occasionally undergo ectopic self-renewal to generate neuroblast/neuroblast siblings; and *pins* mutant neuroblasts occasionally fail to self-renew, resulting in GMC/GMC siblings and termination of the lineage.

We next examined whether *lgl pins* double mutants had fewer neuroblasts (like *pins* mutants) or extra neuroblasts (like *lgl* mutants). Unexpectedly, we detected a phenotype in which the larval brain was full of cells expressing the neuroblast markers Worniu, Miranda and Deadpan and lacking expression of the neuronal marker Elav (Fig. 2b and Supplementary Movie 4). We assayed additional markers that distinguish neuroblasts and GMCs to determine whether these cells were neuroblasts or a hybrid neuroblast/GMC identity. We found that both wild-type neuroblasts and *lgl pins* cells actively transcribed the *worniu*, *deadpan*, *miranda* and *prospero*

genes, maintained proliferation, did not express the Elav neuronal differentiation marker, and did not extend axons (Fig. 2b and Supplementary Fig. 3). The only potential GMC attribute found in *lgl pins* neuroblasts was nuclear Prospero protein (data not shown) but, because wild-type neuroblasts and GMCs both contain Prospero protein, which can accumulate in neuroblast nuclei if not properly localized^{12,13}, this protein is not a definitive marker for the GMC cell type. Thus, *lgl pins* brains contain large numbers of ectopic, proliferating, self-renewing neuroblasts. Combining these *lgl*, *pins* and *lgl pins* mutant data enables us to conclude that *Lgl* inhibits self-renewal, whereas *Pins* has dual functions in promoting and inhibiting self-renewal.

To understand how *Lgl* and *Pins* regulate neuroblast self-renewal at the cellular level, we assayed cortical polarity marker localization in mitotic larval neuroblasts. In wild-type larval neuroblasts, the Par complex (Bazooka–Par6–aPKC) and *Pins*–*Gai* proteins formed an apical crescent at metaphase and were partitioned into the self-renewing neuroblast at telophase, whereas the *Miranda* and *Prospero* proteins formed a basal crescent at metaphase and were partitioned into the differentiating GMC at telophase¹ (Fig. 3a, see legend). In *lgl pins* double mutants, in which all neuroblasts divided symmetrically to generate self-renewing neuroblast/neuroblast siblings, most mitotic neuroblasts showed uniform cortical aPKC, cytoplasmic Bazooka and Par6, and uniform cortical *Miranda* at metaphase and telophase (Fig. 3b, see legend). Thus, only aPKC maintained its correct subcellular localization and correlated with neuroblast self-renewal.

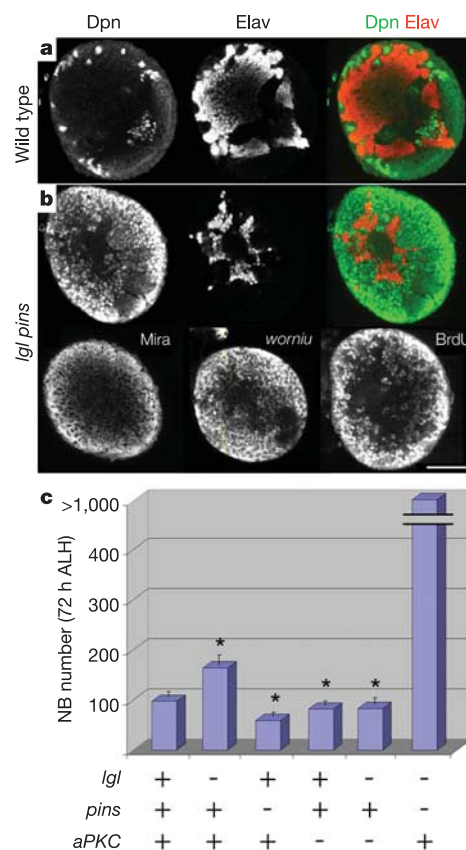


Figure 2 | *lgl pins* double mutants have ectopic neuroblasts and fewer neurons. **a, b**, Wild-type (**a**) and *lgl*³³⁴ *pins*⁶² (**b**) brains at 96 h ALH stained for the neuroblast markers Deadpan, Miranda and *worniu* mRNA, for BrdU incorporation, and for the neuron marker Elav. **c**, Quantification of *lgl*³³⁴, *pins*⁶², and *aPKC*^{K06403} single- and double-mutant phenotypes in brains at 72 h ALH (+, wild type at locus; –, homozygous mutant at locus). Asterisks indicate a statistically significant difference in neuroblast number as compared with wild type ($P \leq 2 \times 10^{-4}$; Student's *t*-test). Scale bar, 50 μ m.

We next examined aPKC localization in *lgl* and *pins* single mutants, in which symmetric divisions occurred at lower frequency. In *lgl* mutants, aPKC showed weak ectopic cortical localization in about half the metaphase neuroblasts, whereas Miranda was delocalized from the cortex; by telophase, however, both proteins appeared to be localized normally^{7,8} (Fig. 3c, see legend). Ectopic cortical aPKC was also observed in *dlg* mutant larval neuroblasts (data not shown). A role for Lgl in restricting aPKC to the apical cortex of neuroblasts has not been reported^{7,8} but would be consistent with the observation that basolateral Lgl restricts aPKC to the apical surface of *Drosophila* and vertebrate epithelia^{14–16} and *Xenopus* blastomeres¹⁷. In *pins* mutants, aPKC and cytoplasmic Miranda showed weak uniform cortical in metaphase neuroblasts, but were properly localized in most telophase neuroblasts (Fig. 3d, see legend). Thus, both Lgl and Pins are required to restrict aPKC to the apical cortex in metaphase neuroblasts.

We next tested whether aPKC is required for neuroblast self-

renewal. *aPKC* mutant clones in larval mushroom body neuroblasts showed premature lineage termination¹⁸, consistent with aPKC being required for neuroblast self-renewal. In addition, *aPKC* null mutants died as second instar larvae with reduced neuroblast numbers (Fig. 2c). Because this was a relatively mild phenotype and there was no detectable aPKC protein at this stage, it is likely that there are additional pathways for stimulating neuroblast self-renewal. We next tested whether aPKC is required for ectopic neuroblast self-renewal in the *lgl* mutants. *lgl aPKC* double mutants had normal numbers of neuroblasts (Fig. 2c), showing that aPKC is required for the ectopic neuroblast self-renewal seen in *lgl* mutants. *aPKC* mutants also suppressed ectopic neuroblast self-renewal in several independently isolated *lgl* mutations, further supporting a role for aPKC in self-renewal (Supplementary Fig. 4). In addition, we found that *aPKC* is fully epistatic to *lgl* in regulating Miranda localization (Fig. 3c, e, f). Thus, aPKC is required for the ectopic neuroblast self-renewal and Miranda delocalization phenotypes seen in *lgl* mutants.

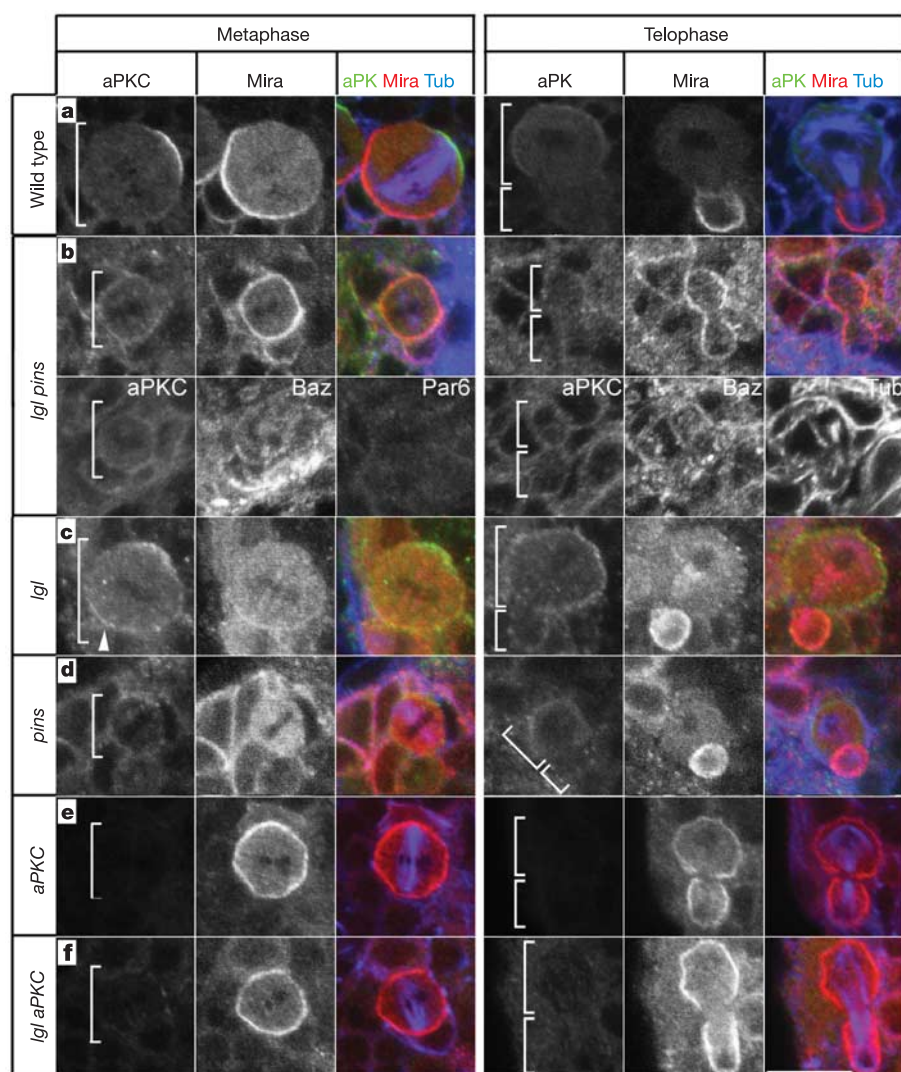


Figure 3 | Lgl and Pins regulate aPKC localization in larval neuroblasts. **a–f**, Brains at 72 h ALH triple-labelled with the indicated markers. Brackets indicate neuroblasts or GMCs; arrowhead indicates ectopic aPKC. **a**, Wild type. aPKC and Miranda form cortical crescents (100%; metaphase, $n = 139$; telophase, $n = 103$). **b**, *lgl³³⁴ pins⁶²* double mutants. At metaphase, aPKC shows ectopic cortical localization (56%; $n = 61$) and Miranda is cortical or cytoplasmic. At telophase, aPKC is cytoplasmic, Miranda is uniform cortical (100% for both; $n = 16$), and Bazooka and Par6 are cytoplasmic. **c**, *lgl³³⁴* mutants. At metaphase, aPKC shows weak ectopic

cortical localization (44%; $n = 93$) and Miranda is delocalized (93%; $n = 93$). At telophase, both are localized normally (100%; $n = 122$). **d**, *pins⁶²* mutants. At metaphase aPKC is uniform cortical and Miranda is cytoplasmic (92% for both; $n = 92$), but by telophase aPKC is apical and Miranda is basal (90% for both; $n = 30$). **e**, *aPKC^{k06403}* mutants. aPKC is undetectable and Miranda is cortical from metaphase to telophase. **f**, *lgl³³⁴ aPKC^{k06403}* double mutants. aPKC is undetectable and Miranda is cortical from metaphase to telophase (100% $n = 36$). Scale bar, 10 μ m.

Our data are most consistent with a model in which Lgl negatively regulates aPKC, and aPKC directly promotes self-renewal. This model is based on the observations that Lgl restricts aPKC localization to the apical cortex of neuroblasts and that a reduction in aPKC blocks the *lgl* self-renewal phenotype. To test this model, we used *worniu-Gal4* line to drive neuroblast-specific expression of constitutively active aPKC or Lgl proteins, and assayed for an increase or decrease in neuroblast numbers. Neuroblast-specific expression of aPKC targeted to the plasma membrane with a CAAX prenylation motif (*UAS-aPKC^{CAAXWT}*)¹⁹ resulted in ectopic cortical aPKC localization, loss of cortical Miranda (Fig. 4a), and a large increase in the number of neuroblasts (Fig. 4g, j, and Supplementary Movie 5). These effects were not observed after overexpression of wild-type aPKC¹⁰ or a membrane-targeted kinase-dead aPKC (*UAS-aPKC^{CAAXKD}*)¹⁹ (Fig. 4b, d, h, j, and Supplementary Movie 6). Expression of a constitutively active aPKC (*UAS-aPKC^{ΔN}*)¹⁰ that was predominantly cytoplasmic (Fig. 4c) gave only a slight increase in neuroblast number (Fig. 4j), showing that cortical localization of aPKC is essential to generate ectopic neuroblasts. By contrast, neuroblast-specific expression of a constitutively active Lgl protein (Lgl3A) resulted in the expected uniform cortical localization of Miranda (Fig. 4e)¹⁰, but no change in neuroblast numbers (Fig. 4j). Combined overexpression of both Lgl3A and aPKC^{CAAXWT}, however, resulted in strong suppression of the aPKC^{CAAXWT} ectopic neuroblast phenotype (Fig. 4i, j), even though Lgl3A alone had no effect on

neuroblast numbers, consistent with Lgl inhibiting aPKC function either directly or at through its downstream effectors. Thus, neuroblast-specific overexpression of aPKC can expand the neuroblast population (most probably by promoting symmetric neuroblast/neuroblast cell divisions) without eliminating the ability of these neuroblasts to undergo asymmetric neuroblast/GMC divisions to generate differentiating progeny. We conclude that aPKC is sufficient to promote neuroblast self-renewal, Lgl can inhibit aPKC function, and membrane-targeting and kinase activity are essential for aPKC function.

We have established *Drosophila* larval neuroblasts as a model system for studying self-renewal versus differentiation. We have proposed a simple model in which Pins anchors aPKC apically and Lgl inhibits aPKC localization basally, thereby restricting aPKC to the apical cortex where it promotes neuroblast self-renewal (Fig. 4k). In addition, aPKC can phosphorylate and directly inhibit Lgl function¹⁰, which together with our data provides evidence for mutual inhibition between Lgl and aPKC in neuroblasts, similar to the mutual inhibition seen between these two proteins in epithelia^{14–17}. Mutual inhibition between aPKC and Lgl would result in stabilization of apical aPKC localization and more reliable partitioning of aPKC into the neuroblast during mitosis. In *pins* mutants, aPKC is delocalized and nonfunctional owing to Lgl activity, thereby reducing self-renewal; in *lgl* mutants, aPKC shows weak ectopic cortical localization that increases self-renewal; and in *lgl pins* double mutants, aPKC is both delocalized and fully active, and thus all neuroblasts undergo symmetric self-renewal (Fig. 4k). Although the targets of aPKC involved in self-renewal are unknown, aPKC may directly phosphorylate and inactivate GMC determinants²⁰, and/or phosphorylate and activate neuroblast-specific proteins. Notably, *lgl* mutant mice have neural progenitor hypertrophy²¹ and knockdown of a *pins* mammalian homologue (AGS3) leads to depletion of neural progenitors²²; phenotypes that are very similar to those described here. In the future, it will be important to determine the role of aPKC in mammalian neural progenitor self-renewal and to identify the aPKC-regulated phosphoproteins that regulate neuroblast self-renewal in *Drosophila*.

METHODS

All fly stocks have been described^{10,11,18,19,23,24}. Neuroblast clones were generated as described²⁵ using a 90-min heat shock at 37 °C to larvae at 24 h ALH with recovery at 25 °C. Published methods were used for BrdU pulse labelling⁴, *worniu* ribonucleotide probe generation²⁶, fluorescent *in situ* hybridization²⁷ and antibody staining¹⁸ (details are available on request from the authors and in the Supplementary Methods).

Received 27 July; accepted 3 October 2005.

Published online 14 December 2005.

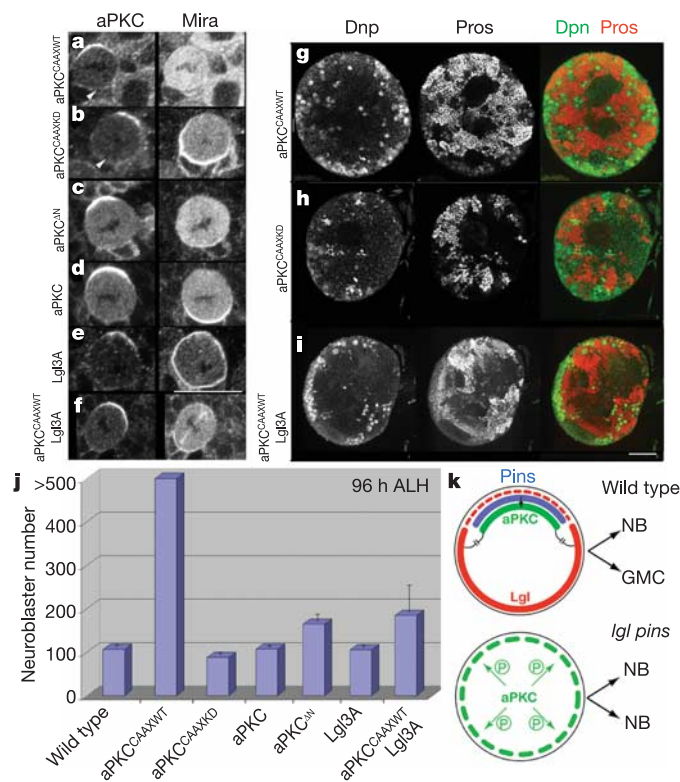


Figure 4 | Overexpression of cortical aPKC promotes neuroblast self-renewal. **a–i**, Expression of the indicated *UAS* transgenes (left) by *worniu-Gal4* in brains at 96 h ALH stained for the indicated markers (top). All neuroblasts expressed the correct molecular markers (Deadpan⁺, nuclear-Prospero⁺), were proliferative (BrdU⁺) and could even generate smaller nuclear-Prospero⁺ GMCs (**g–i**; data not shown). Scale bars, 10 μm (**a–f**); 50 μm (**g–i**). Arrowheads indicate ectopic aPKC; asterisks indicate aPKC⁺ GMCs contacting neuroblast. **j**, Quantification of neuroblast number (Student *t*-test). Wild type, 98.9 ± 4.6; aPKC^{CAAXWT}, >500; aPKC^{CAAXKD}, 87.4 ± 4.6; aPKC, 99.5 ± 3; aPKC^{ΔN}, 150.4 ± 15.7; Lgl3A, 97.6 ± 4.4; aPKC^{CAAXWT} Lgl3A, 173.6 ± 61.7 (means ± s.d.). **k**, Model of the role of Pins, Lgl and aPKC in regulating neuroblast self-renewal. See text for details.

- Betschinger, J. & Knoblich, J. A. Dare to be different: asymmetric cell division in *Drosophila*, *C. elegans* and vertebrates. *Curr. Biol.* **14**, R674–R685 (2004).
- Ohlstein, B., Kai, T., Decotto, E. & Spradling, A. The stem cell niche: theme and variations. *Curr. Opin. Cell Biol.* **16**, 693–699 (2004).
- Urbach, R. & Technau, G. M. Neuroblast formation and patterning during early brain development in *Drosophila*. *BioEssays* **26**, 739–751 (2004).
- Datta, S. Control of proliferation activation in quiescent neuroblasts of the *Drosophila* central nervous system. *Development* **121**, 1173–1182 (1995).
- Gateff, E. & Schneiderman, H. A. Developmental capacities of benign and malignant neoplasms of *Drosophila*. *Roux Arch. Dev. Biol.* **176**, 23–65 (1974).
- Ohshiro, T., Yagami, T., Zhang, C. & Matsuzaki, F. Role of cortical tumour-suppressor proteins in asymmetric division of *Drosophila* neuroblast. *Nature* **408**, 593–596 (2000).
- Peng, C. Y., Manning, L., Albertson, R. & Doe, C. Q. The tumour-suppressor genes *lgl* and *dlg* regulate basal protein targeting in *Drosophila* neuroblasts. *Nature* **408**, 596–600 (2000).
- Albertson, R. & Doe, C. Q. *Dlg*, *Scrib* and *Lgl* regulate neuroblast cell size and mitotic spindle asymmetry. *Nature Cell Biol.* **5**, 166–170 (2003).
- Strand, D. *et al.* The *Drosophila* lethal(2)giant larvae tumour suppressor protein forms homo-oligomers and is associated with nonmuscle myosin II heavy chain. *J. Cell Biol.* **127**, 1361–1373 (1994).
- Betschinger, J., Mechtler, K. & Knoblich, J. A. The Par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein Lgl. *Nature* **422**, 326–330 (2003).

11. Buenzow, D. E. & Holmgren, R. Expression of the *Drosophila* gooseberry locus defines a subset of neuroblast lineages in the central nervous system. *Dev. Biol.* **170**, 338–349 (1995).
12. Freeman, M. R. & Doe, C. Q. Asymmetric Prospero localization is required to generate mixed neuronal/glial lineages in the *Drosophila* CNS. *Development* **128**, 4103–4112 (2001).
13. Spana, E. P. & Doe, C. Q. The Prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. *Development* **121**, 3187–3195 (1995).
14. Bilder, D., Li, M. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumour suppressors. *Science* **289**, 113–116 (2000).
15. Hutterer, A., Betschinger, J., Petronczki, M. & Knoblich, J. A. Sequential roles of Cdc42, Par-6, aPKC, and Lgl in the establishment of epithelial polarity during *Drosophila* embryogenesis. *Dev. Cell* **6**, 845–854 (2004).
16. Yamanaka, T. *et al.* Mammalian Lgl forms a protein complex with PAR-6 and aPKC independently of PAR-3 to regulate epithelial cell polarity. *Curr. Biol.* **13**, 734–743 (2003).
17. Chalmers, A. D. *et al.* aPKC, Crumbs3 and Lgl2 control apicobasal polarity in early vertebrate development. *Development* **132**, 977–986 (2005).
18. Rolls, M. M., Albertson, R., Shih, H. P., Lee, C. Y. & Doe, C. Q. *Drosophila* aPKC regulates cell polarity and cell proliferation in neuroblasts and epithelia. *J. Cell Biol.* **163**, 1089–1098 (2003).
19. Sotillos, S., Diaz-Meco, M. T., Caminero, E., Moscat, J. & Campuzano, S. DaPKC-dependent phosphorylation of Crumbs is required for epithelial cell polarity in *Drosophila*. *J. Cell Biol.* **166**, 549–557 (2004).
20. Caussinus, E. & Gonzalez, C. Induction of tumour growth by altered stem-cell asymmetric division in *Drosophila melanogaster*. *Nature Genet.* **37**, 1125–1129 (2005).
21. Klezovitch, O., Fernandez, T. E., Tapscott, S. J. & Vasioukhin, V. Loss of cell polarity causes severe brain dysplasia in *Lgl1* knockout mice. *Genes Dev.* **18**, 559–571 (2004).
22. Sanada, K. & Tsai, L. H. G Protein $\beta\gamma$ subunits and AGS3 control spindle orientation and asymmetric cell fate of cerebral cortical progenitors. *Cell* **122**, 119–131 (2005).
23. Yu, F., Morin, X., Cai, Y., Yang, X. & Chia, W. Analysis of partner of inscuteable, a novel player of *Drosophila* asymmetric divisions, reveals two distinct steps in inscuteable apical localization. *Cell* **100**, 399–409 (2000).
24. Albertson, R., Chabu, C., Sheehan, A. & Doe, C. Q. Scribble protein domain mapping reveals a multistep localization mechanism and domains necessary for establishing cortical polarity. *J. Cell Sci.* **117**, 6061–6070 (2004).
25. Pearson, B. J. & Doe, C. Q. Regulation of neuroblast competence in *Drosophila*. *Nature* **425**, 624–628 (2003).
26. Freeman, M. R., Delrow, J., Kim, J., Johnson, E. & Doe, C. Q. Unwrapping glial biology: Gcm target genes regulating glial development, diversification, and function. *Neuron* **38**, 567–580 (2003).
27. Grosskortenhaus, R., Pearson, B. J., Marusich, A. & Doe, C. Q. Regulation of temporal identity transitions in *Drosophila* neuroblasts. *Dev. Cell* **8**, 193–202 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Knoblich, S. Campuzano, B. Chia, J. Skeath and B. Holmgren for fly stocks and/or antibody reagents; B. Bowerman, J. Eisen, K. Siller and S. Siegrist for comments on the manuscript; and C. Chabu for discussion. C.-Y.L. is supported by a Damon Runyon postdoctoral fellowship. C.Q.D. is supported by the Howard Hughes Medical Institute, where he is an Investigator.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.Q.D. (cdoe@uoregon.edu).

Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor

Gregor B. Adams^{1,3}, Karissa T. Chabner^{1,3}, Ian R. Alley^{1,3}, Douglas P. Olson¹, Zbigniew M. Szczepiorkowski^{1,2,†}, Mark C. Poznansky¹, Claudine H. Kos⁴, Martin R. Pollak⁴, Edward M. Brown⁵ & David T. Scadden^{1,3}

During mammalian ontogeny, haematopoietic stem cells (HSCs) translocate from the fetal liver to the bone marrow, where haematopoiesis occurs throughout adulthood¹. Unique features of bone that contribute to a microenvironmental niche for stem cells might include the known high concentration of calcium ions at the HSC-enriched endosteal surface. Cells respond to extracellular ionic calcium concentrations through the seven-transmembrane-spanning calcium-sensing receptor (CaR), which we identified as being expressed on HSCs. Here we show that, through the CaR, the simple ionic mineral content of the niche may dictate the preferential localization of adult mammalian haematopoiesis in bone. Antenatal mice deficient in CaR had primitive haematopoietic cells in the circulation and spleen, whereas few were found in bone marrow. CaR^{-/-} HSCs from fetal liver were normal in number, in proliferative and differentiative function, and in migration and homing to the bone marrow. Yet they were highly defective in localizing anatomically to the endosteal niche, behaviour that correlated with defective adhesion to the extracellular matrix protein, collagen I. CaR has a function in retaining HSCs in close physical proximity to the endosteal surface and the regulatory niche components associated with it.

Under homeostatic conditions, HSCs within the bone marrow (BM) cavity reside very close to the endosteal surfaces of bone^{2,3}, and transplanted HSCs migrate to these areas within hours of intravenous injection⁴. Endochondral ossification has been shown to be an essential prerequisite for the development of normal haematopoiesis in the BM⁵⁻⁷, indicating a possible fundamental interrelationship of ossification to the mature haematopoietic process in mammals. In support of this hypothesis, recent reports by us and others have identified that a key cellular component of the HSC niche is cells of the osteoblast lineage, the cell type responsible for the formation of bone^{8,9}.

A characteristic of the endosteum where active bone modelling and remodelling takes place is an increased extracellular calcium-ion concentration [Ca²⁺]_o, which reaches 40 mM near resorbing osteoclasts¹⁰, more than 20-fold that observed in serum. One mechanism by which cells can respond to variations in [Ca²⁺]_o is through the CaR, a G-protein-coupled receptor first cloned from bovine parathyroid and known to participate in parathyroid hormone response to [Ca²⁺]_o levels¹¹. Expression of the CaR has also been shown on haematopoietic cells, including cells of the monocyte/macrophage lineage in which activation induces their transmigration *in vitro* and *in vivo*^{12,13}. We reasoned that, if expressed on HSCs, CaR might affect stem cell responses to the unique endosteal microenvironment.

Cell-surface CaR was found by flow cytometry on BM-derived, stem-cell-enriched, lineage-negative (lin⁻) Sca-1⁺c-Kit⁺ (LSK) cells from wild-type mice (Fig. 1a) and expression was confirmed by quantitative real-time reverse transcriptase-mediated polymerase chain reaction (RT-PCR) (Supplementary Fig. 1). Histological analysis was then performed on the BM of CaR^{+/+} or CaR^{-/-} animals^{14,15} using mice within 72 h of birth, because beyond that age CaR^{-/-} animals become hypercalcaemic and die by the age of 7–10 days. Marked hypocellularity was observed in the BM of the CaR^{-/-} mice relative to littermate controls (Fig. 1b). The apparent size of the BM cavity was minimally changed and trabecular bone was intact, but the marrow space contained a disproportionately small number of haematopoietic cells with paratrabecular haematopoietic clusters notably absent (Fig. 1b). These observations were confirmed by direct measurement of mononuclear cell (MNC) counts of femur marrow obtained by crushing the bones (Fig. 1c). Flow cytometric analyses on BM MNCs demonstrated similar relative frequencies of T cells, B cells, granulocytes and monocytes in CaR^{-/-} and CaR^{+/+} mice (Fig. 1d), indicating that there was not an abnormality of differentiation. The single cell type that was disproportionately decreased was those expressing the erythrocyte lineage marker Ter119, a subpopulation known to express CaR but of unknown function.

To address the question of whether the reduced cellularity in the BM of the CaR^{-/-} mice resulted from a lack of primitive haematopoietic elements, we examined the number of HSCs and haematopoietic progenitor cells (HPCs) with the use of both phenotypic and functional assays. Primitive LSK cells in CaR^{-/-} animals were reduced 2.5-fold in relative frequency (Fig. 2a; $P = 0.016$) and about sevenfold in absolute number (4,547 in CaR^{-/-} animals versus 33,352 in CaR^{+/+} animals) compared with wild-type littermates. Functionally, long-term culture-initiating cells (LTC-ICs) assessed at limiting dilution, an *in vitro* surrogate for HSCs¹⁶, were also decreased in mean frequency (9.95 per 10⁵ BM MNCs in CaR^{-/-} animals versus 18.22 in CaR^{+/+} animals; $P = 0.0057$) and absolute number in the CaR^{-/-} mice (Fig. 2b). Although more mature HPCs, assessed by the methylcellulose-based colony-forming unit culture (CFU-C) assay, were increased in relative mean frequency (50.5 per 10⁴ BM MNCs in CaR^{-/-} animals versus 32 in CaR^{+/+} animals; $P = 0.0025$), the total numbers of CFU-Cs per femur were decreased (Fig. 2c). The somewhat differing profiles of stem versus progenitor populations in the BM may reflect distinct regulatory mechanisms governing specific cell 'compartments' in haematopoiesis, a concept supported by substantial experimental evidence from us and others¹⁷⁻¹⁹.

To assess whether reduced HSCs in BM was due to aberrant

¹Center for Regenerative Medicine and ²Blood Transfusion Service, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. ³Harvard Stem Cell Institute, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Renal Division and ⁵Endocrine-Hypertension Division, Membrane Biology Program, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [†]Present address: Department of Pathology, Dartmouth-Hitchcock Medical Center, Dartmouth Medical School, Lebanon, New Hampshire 03756, USA.

localization, we analysed the spleen and blood of $\text{CaR}^{+/+}$ and $\text{CaR}^{-/-}$ animals. Absolute cellularity of the spleen was not significantly different (38.3×10^6 MNCs in $\text{CaR}^{-/-}$ animals versus 39×10^6 MNCs in $\text{CaR}^{+/+}$ animals; $P = 0.964$); however, $\text{CaR}^{-/-}$ animals had a mean twofold increase in the frequency of both splenic LSK cells (Fig. 2a; $P = 0.027$) and CFU-C (Fig. 2d). Similarly, LSK cells were increased fivefold in the blood of $\text{CaR}^{-/-}$ mice (Fig. 2a; $P = 0.012$). Primitive populations of cells are therefore resident in sites other than the BM in the absence of CaR.

To assess whether the stem-cell defect was due to an intrinsic abnormality of the HSCs, we examined primitive haematopoietic populations before the transmigration of haematopoiesis from fetal liver (FL) to BM. Quantifying embryonic-day 17 (E17) FL HSCs, we observed comparable mean frequencies by both phenotypic (LSK)

(Fig. 3a; $P = 0.21$) and functional (LTC-IC) assays (Fig. 3b); comparable progenitor colonies emerged after the long-term culture, indicating intact differentiation. Owing to the known effects of cell cycling on HSC engraftment²⁰, cell cycle profiles of LSK were examined and found to be indistinguishable (Fig. 3c). Reduced primitive populations in the BM of $\text{CaR}^{-/-}$ animals could therefore not be attributed to an aberrant production or function of HSCs before translocation.

To assess whether there was an intrinsic engraftment defect in the $\text{CaR}^{-/-}$ cells specific to the BM niche, we transplanted FL MNCs into lethally irradiated wild-type mice. FL MNCs from either $\text{CaR}^{+/+}$ or $\text{CaR}^{-/-}$ littermates gave 100% survival. However, the BM of the mice transplanted with $\text{CaR}^{-/-}$ cells had markedly fewer resident HSCs and HPCs (Fig. 3d, e). It is not clear why cellularity was comparable between genotypes after transplantation whereas cellularity was decreased in $\text{CaR}^{-/-}$ under homeostatic conditions, but differences in the post-transplant context such as the cytokine milieu probably contribute. Although we did not observe cell cycle differences in the FL cells, we reasoned that altered HSC numbers might be due to differential cell cycling after transplantation; however, we could detect no differences in lin^- cells in G0 ($P = 0.927$) or G1 ($P = 0.921$) after transplant (Supplementary Fig. 2). Competitive repopulation experiments were then performed with the congenic CD45.1/CD45.2 allele to distinguish donor and competitor cells. CD45.2 $\text{CaR}^{-/-}$ FL MNCs were transplanted in equal amount with CD45.1 competitor wild-type BM MNCs, yet they yielded only 2.7% of BM haematopoietic cells in the recipient 16 weeks after transplantation. In contrast, $\text{CaR}^{+/+}$ CD45.2 FL MNCs competing with wild-type CD45.1 BM MNCs contributed to 84.6% of BM haematopoietic cells ($P = 0.0002$; Fig. 3f). $\text{CaR}^{-/-}$ cells therefore engraft poorly in a setting where there exists stem cell competition for the BM endosteal niche. These results indicate that the BM haematopoietic

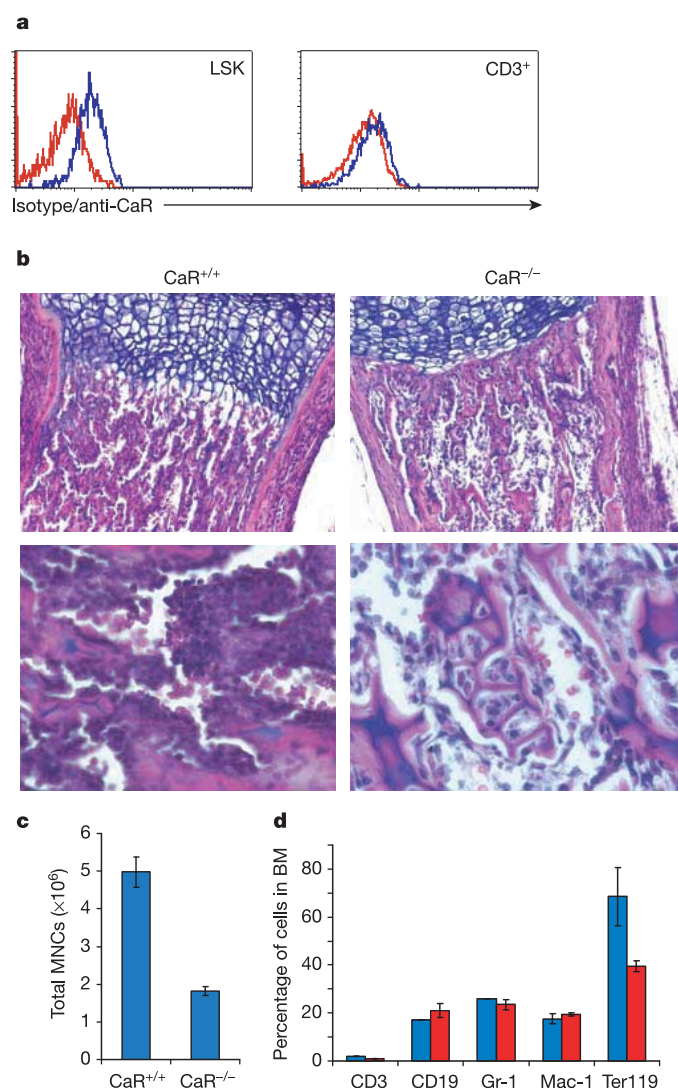


Figure 1 | $\text{CaR}^{-/-}$ mice have a hypocellular bone marrow. **a**, The expression of the CaR on the cell surface of primitive BM LSK cells was verified by flow cytometry using an anti-CaR antibody. The expression on CD3^+ T-cells, of which only a minority expresses CaR, is also shown. Isotype control is shown in red, anti-CaR staining in blue. **b**, Femurs were isolated from 3-day-old $\text{CaR}^{+/+}$ and $\text{CaR}^{-/-}$ mice and assessed for cellularity in the BM cavity by staining with haematoxylin and eosin. Top, original magnification $\times 100$; bottom, $\times 400$. **c**, Total MNCs obtained from four limbs of 3-day-old mice ($n = 4$ in each group). The result is significant ($P = 0.001$). **d**, The relative proportions of the different haematopoietic cell lineages were measured by flow cytometry ($n = 4$ in each group). Blue, $\text{CaR}^{+/+}$; red, $\text{CaR}^{-/-}$. Error bars show s.e.m.

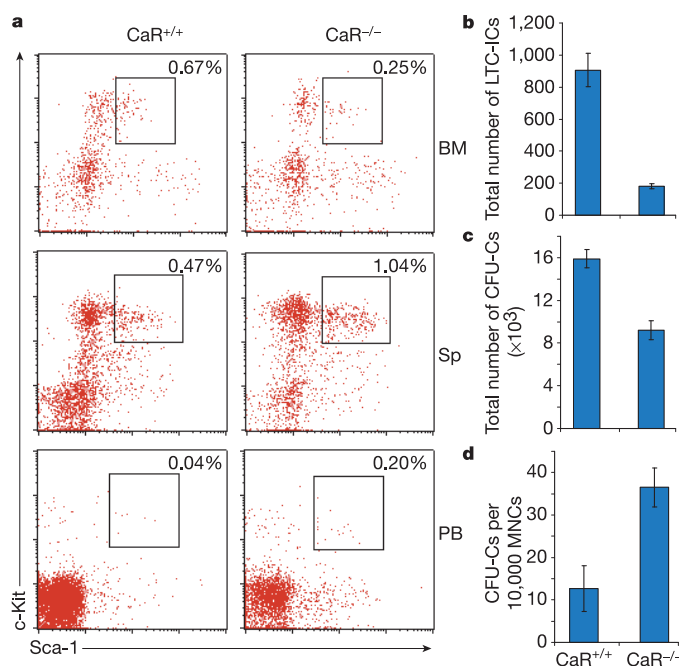


Figure 2 | $\text{CaR}^{-/-}$ mice have aberrant localization of primitive haematopoietic cells. **a**, MNCs from BM, spleen (Sp) and peripheral blood (PB) of 3-day-old $\text{CaR}^{+/+}$ and $\text{CaR}^{-/-}$ mice were assessed for the frequency of LSK cells. The expression of Sca-1 and c-Kit on lin^- cells is shown. **b**, **c**, The total number of LTC-ICs (**b**) and CFU-Cs (**c**) in the BM of $\text{CaR}^{+/+}$ and $\text{CaR}^{-/-}$ mice. Significance of differences: $P = 0.006$ (**b**); $P = 0.003$ (**c**). **d**, The frequency of CFU-Cs in the MNC fraction of the spleen of $\text{CaR}^{+/+}$ and $\text{CaR}^{-/-}$ mice ($n = 5$ in each group). The result is significant at the $P = 0.004$ level. Error bars show s.e.m.

defects observed in the $\text{CaR}^{-/-}$ mice are stem cell autonomous and are specific to engraftment in the BM microenvironment. They do not depend on the $\text{CaR}^{-/-}$ host and possible effects from other abnormalities in the mouse, such as hyperparathyroidism or hypercalcaemia.

This observed altered engraftment of HSCs in the BM may be due to altered homing or retention of the cells in the correct microenvironment. To examine this, we assessed on FL LSK cells the surface expression of molecules believed to be involved in the homing of HSCs to the BM, including CD49d, CD62L and CXCR4; however, we found no alteration between $\text{CaR}^{+/+}$ or $\text{CaR}^{-/-}$ cells in any of these attachment or migration mediators (Fig. 4a). Assessing migratory function, no defect was observed in $\text{CaR}^{-/-}$ cell response to a gradient of stromal cell-derived factor-1 α (SDF-1 α) (Fig. 4b). Further, *in vivo* homing assays of fluorescently labelled primitive FL MNCs showed no differences in the number of $\text{CaR}^{+/+}$ or $\text{CaR}^{-/-}$ cells seeding the marrow space (Fig. 4c). However, engraftment also requires HSC engagement and retention in the

correct microenvironmental niche, a process that has been termed 'lodgment'²¹. The basis for this is probably downstream effects of CaR activation. Signalling through CaR is known to involve $G_{i\alpha}$ and $G_{q\alpha}$ with subsequent inhibition of cyclic AMP generation and activation of mitogen-activated protein kinase (MAPK) cascades and phospholipase C, respectively. We detected no changes in intracellular MAPK or extracellular signal-related kinase 1 phosphorylation in LSK cells after stimulation of the CaR (data not shown), although we previously noted an intracellular calcium shift in haematopoietic cells after stimulation with either ionic calcium or allosteric agonist¹². To determine whether CaR depletion altered the surface expression of N-cadherin, a potential mediator of stem cell–niche interactions⁹, we assessed $\text{CaR}^{+/+}$ and $\text{CaR}^{-/-}$ LSK cells and detected no difference (Fig. 4a). Next we assessed the adherence of the

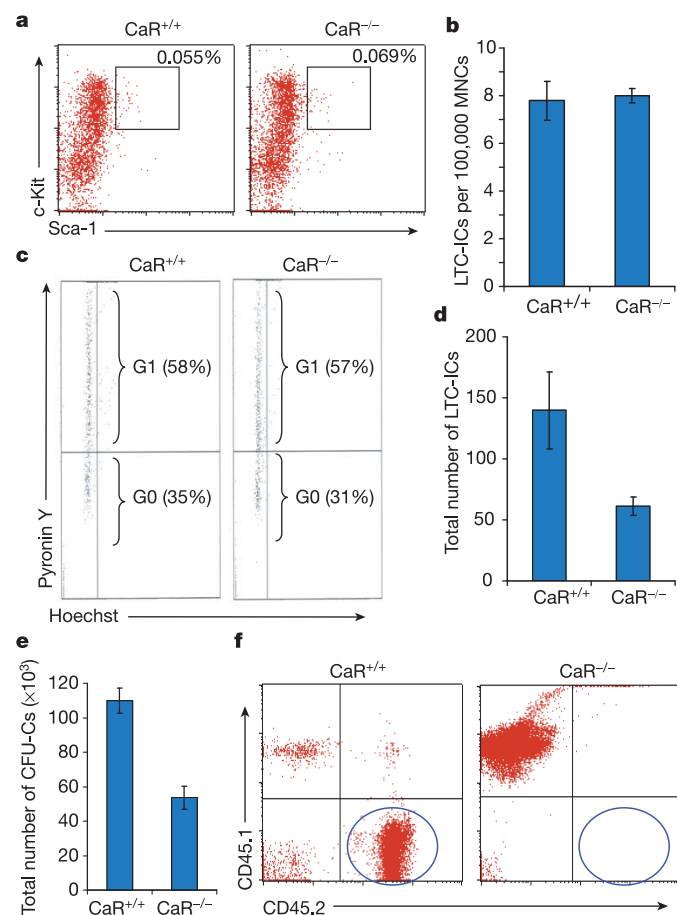


Figure 3 | $\text{CaR}^{-/-}$ HSC defects are stem cell autonomous, specific to the bone marrow microenvironmental niche. **a**, **b**, MNC suspensions were obtained from the FL of E17.5 embryos and the numbers of HSCs were assessed by both Sca-1 and c-Kit expression on lin^{-} cells (**a**) and the functional LTC-IC assay (**b**). The result in **b** is not significant ($P = 0.675$). **c**, Analysis of the proportion of the FL lin^{-} cells in the G0 versus G1 phase of the cell cycle is shown ($n = 3$ in each group). **d**, **e**, MNCs from $\text{CaR}^{+/+}$ or $\text{CaR}^{-/-}$ E17.5 FL were injected into wild-type hosts. LTC-IC number (**d**) and CFU-C number (**e**) were assessed 16 weeks after transplantation. Significance of differences: $P = 0.027$ (**d**); $P = 0.0008$ (**e**). **f**, Analysis of the competitive transplantation of CD45.2 FL cells and CD45.1 wild-type BM cells into CD45.1 hosts. The contribution to haematopoiesis from the FL cells (CD45.2⁺) is highlighted ($n = 5$ in each group). Error bars show s.e.m.

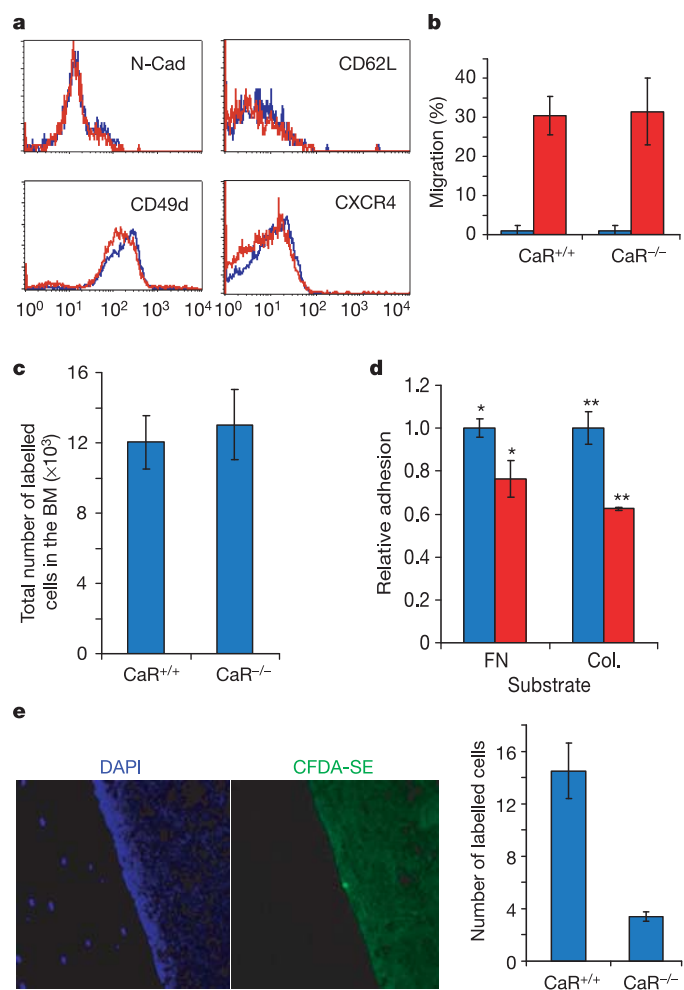


Figure 4 | $\text{CaR}^{-/-}$ HSCs are unable to localize to the endosteal niche. **a**, Expression of adhesion molecules N-cadherin (N-Cad), L-selectin (CD62L), VLA-4 (CD49d) and CXCR4 on the surface of FL LSK cells. Blue, $\text{CaR}^{+/+}$; red, $\text{CaR}^{-/-}$. **b**, *In vitro* chemotaxis of FL lin^{-} cells to 300 ng ml⁻¹ SDF-1 α . Blue, chemokinesis; red, chemotaxis. The result is not significant ($P = 0.899$). **c**, Homing of CFDA-SE-labelled lin^{-} cells from FL to the BM of wild-type recipients. The result is not significant ($P = 0.653$). **d**, Adhesion of lin^{-} cells from E17.5 FLs to plates coated with fibronectin (FN) and collagen I (Col.). The relative adhesion of the cells compared to that of $\text{CaR}^{+/+}$ cells is shown. Blue, $\text{CaR}^{+/+}$; red, $\text{CaR}^{-/-}$. Asterisk, $P = 0.073$; two asterisks, $P = 0.02$. **e**, Quantification of the number of lin^{-} cells labelled with CFDA-SE within two cell diameters of the endosteal surface in 20 tibial sections in wild-type hosts. An example of the specific localization of a $\text{CaR}^{+/+}$ cell at the endosteal niche of the BM is shown. DAPI, 4',6-diamidino-2-phenylindole. The result is significant ($P = 0.018$). Error bars show s.e.m.

primitive FL haematopoietic cells to substrates found in the BM microenvironment, where the HSCs are also located, specifically fibronectin and collagen I²². We found a significant defect in the ability of the CaR^{-/-} cells to adhere to collagen I, which is the most prominent extracellular matrix protein of bone and is produced by osteoblasts (Fig. 4d). Finally, we examined by direct visualization whether primitive cells could engage the niche *in vivo*. Using histological assessment of transplanted, labelled cells, we scored cells within two cell diameters of the endosteal surface and observed a marked reduction in the ability of the CaR^{-/-} cells to 'lodge' in this microenvironment (Fig. 4e). Impaired physical association of CaR^{-/-} HSCs with the regulatory components of the microenvironmental niche might therefore account for the observed defect in BM HSC engraftment.

These abnormalities found in the CaR^{-/-} HSCs were a cell autonomous defect in their ability to bind to extracellular matrix molecules present at the endosteal surface of bone, specifically collagen I, one of the major matrix molecules released by cells of the osteoblastic lineage. This resulted in a specific defect in the ability of the HSCs to lodge at the endosteal niche, and thus engraft, in the adult BM. However, the results indicate that HSCs might preferentially localize to the cells releasing extracellular calcium, the osteoclasts, rather than to osteoblasts. Osteoblasts and osteoclasts are thought to form a functional unit of bone remodelling with immediate cell proximity. We propose that HSCs join that unit, with local [Ca²⁺]_o altering HSC function directly through CaR and indirectly through the known effects of calcium on select adhesive interactions, particularly N-cadherin known to be at the HSC/osteoblast interface⁹.

The ability of the CaR to specify BM localization for HSCs has two implications. The first is in the movement of HSCs from FL to BM during development, which occurs in mammals during the time of bone mineralization. Before that period, bony structures are cartilaginous and without defined marrow tissue. However, recapitulating this process through ectopic generation of calcified bony tissue by injection of BMP results in localized haematopoietic tissue, again indicating that calcified bone might have a key function in localization of adult haematopoiesis^{23,24}. The second implication of this work is in the area of clinical stem cell transplantation, in which modulating CaR with drugs may be envisioned as a strategy either to enhance the engraftment of transplanted HSCs in the BM niche or, conversely, to mobilize stem cells from it.

METHODS

Flow cytometric analysis. Immunophenotypic enumeration of the HSC frequency and cell cycle analysis was performed as described previously⁸. CaR expression was detected with an anti-CaR rabbit polyclonal antibody (Affinity Bioreagents). For analysis of adhesion molecule expression in the LSK population, MNCs were stained with the anti-lineage, Sca-1 and CD117 antibodies, as well as either anti-CD49d, anti-CXCR4, anti-CD62L or anti-N-cadherin antibodies (Pharminogen).

Q-PCR. RNA was isolated from cells using the Absolutely RNA isolation kit (Stratagene). Complementary DNA was then made from the RNA by using the Promega Reverse Transcription System in accordance with the manufacturer's instructions. To quantify the expression of *car* and *hprt*, the Assays-on-Demand primer and probe sets were used for their respective genes (Applied Biosystems). Levels of expression were quantified with the Mx3000P real-time PCR system (Stratagene). Standard curves in these experiments were obtained with total mouse kidney RNA (Stratagene) subjected to the same cDNA amplification protocol.

Colony-forming unit assay. MNCs isolated from BM or spleen were assessed for CFU-C frequency by culturing in methylcellulose-containing medium M3434 (Stem Cell Technologies) as described previously¹⁸.

LTC-IC assay. LTC-IC frequency in BM and FL MNCs was quantified as described previously²⁵.

Engraftment studies. Mice were obtained and used in accordance with the guidelines of the Subcommittee on Research Animal Care of the Massachusetts General Hospital. Mice were housed in sterilized microisolator cages and received autoclaved food and water *ad libitum*. For the engraftment experiments,

10⁶ FL MNCs were injected into B6.SJL mice (Jackson Laboratories) that had been lethally irradiated (10 Gy) 24 h previously. After 16 weeks the mice were killed with CO₂ and the BM was removed and flushed with fully supplemented Iscove's medium. For the competitive transplantation experiments, 2 × 10⁵ FL MNCs from CaR^{+/+} or CaR^{-/-} mice were mixed with 2 × 10⁵ BM MNCs cells from a B6.SJL mouse, and the relative contributions to haematopoiesis were assessed as described previously⁸.

In vitro transmigration. Chemotaxis assays were performed with a transwell with a pore size of 5 µm (Corning-Costar Corp.). Purified lin⁻ cells (5 × 10⁴) in fully supplemented Iscove's medium were added to the upper well. Chemotaxis towards 300 ng ml⁻¹ murine SDF-1α (PeproTech Inc.) in the lower chamber was allowed to continue for 3 h at 37 °C under 5% CO₂ in a humidified atmosphere. Cells were harvested from the lower well and counted.

Homing in vivo. MNCs were labelled with 5 µM carboxyfluorescein diacetate succinimidyl ester (CFDA-SE) (Molecular Probes Inc.) in accordance with the manufacturer's instructions. Cells were then injected into the tail vein of 6–8-week-old C57BL/6 mice that had been lethally irradiated (10 Gy) 24 h previously. Mice were then killed 6 h after injection and the numbers of homed cells were measured in the BM through the detection of CFDA-SE⁺ cells by flow cytometry. To assess homing of the cells to the endosteal surface, labelled cells were injected into non-irradiated animals as described previously for these analyses⁴. Tibias were dissected and fixed in 10% buffered formalin. The bones were then decalcified for 3–4 days in Immunocal (Decal Corp.) and embedded in paraffin blocks. Sections 6 µm thick were cut, deparaffinized and mounted with Vectashield (Vector Laboratories) containing 4',6-diamidino-2-phenylindole (Molecular Probes). Cells homing to the endosteal niche were identified as being a maximum of two cell diameters from the endosteal surface.

Adhesion in vitro. Isolated lin⁻ cells in fully supplemented medium were plated on plates coated with fibronectin or collagen I (BD Biosciences). Cultures were incubated for 3 h at 37 °C under 5% CO₂ in a humidified atmosphere. Non-adherent cells were removed with five washes with PBS, after which adherent cells were quantified by light microscopy.

Statistical analysis. Results are expressed as means ± s.e.m. Data were analysed with the unpaired two-tailed Student's *t*-test as appropriate for the data set. *P* < 0.05 was considered significant.

Received 8 August; accepted 15 September 2005.

Published online 28 December 2005.

1. Tavassoli, M. Embryonic and fetal hemopoiesis: an overview. *Blood Cells* **17**, 269–281 (1991).
2. Lord, B. I., Testa, N. G. & Hendry, J. H. The relative spatial distributions of CFUs and CFUCs in the normal mouse femur. *Blood* **46**, 65–72 (1975).
3. Gong, J. K. Endosteal marrow: a rich source of hematopoietic stem cells. *Science* **199**, 1443–1445 (1978).
4. Nilsson, S. K., Johnston, H. M. & Coverdale, J. A. Spatial localization of transplanted hematopoietic stem cells: inferences for the localization of stem cell niches. *Blood* **97**, 2293–2299 (2001).
5. Zhou, H. *et al.* Marrow development and its relationship to bone formation *in vivo*: a histological study using an implantable titanium device in rabbits. *Bone* **17**, 407–415 (1995).
6. Deguchi, K. *et al.* Excessive extramedullary hematopoiesis in Cbfa1-deficient mice with a congenital lack of bone marrow. *Biochem. Biophys. Res. Commun.* **255**, 352–359 (1999).
7. Jacenko, O. *et al.* Linking hematopoiesis to endochondral skeletogenesis through analysis of mice transgenic for collagen X. *Am. J. Pathol.* **160**, 2019–2034 (2002).
8. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
9. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **415**, 836–841 (2003).
10. Silver, I. A., Murrills, R. J. & Etherington, D. J. Microelectrode studies on the acid microenvironment beneath adherent macrophages and osteoclasts. *Exp. Cell Res.* **175**, 266–276 (1988).
11. Chattopadhyay, N., Vassilev, P. M. & Brown, E. M. Calcium-sensing receptor: roles in and beyond systemic calcium homeostasis. *Biol. Chem.* **378**, 759–768 (1997).
12. House, M. G. *et al.* Expression of an extracellular calcium-sensing receptor in human and mouse bone marrow cells. *J. Bone Miner. Res.* **12**, 1959–1970 (1997).
13. Olszak, I. T. *et al.* Extracellular calcium elicits a chemokinetic response from monocytes *in vitro* and *in vivo*. *J. Clin. Invest.* **105**, 1299–1305 (2000).
14. Ho, C. *et al.* A mouse model of human familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Nature Genet.* **11**, 389–394 (1995).
15. Garner, S. C., Pi, M., Tu, Q. & Quarles, L. D. Rickets in cation-sensing receptor-deficient mice: an unexpected skeletal phenotype. *Endocrinology* **142**, 3996–4005 (2001).

16. Sutherland, H. J., Lansdorp, P. M., Henkelman, D., Eaves, A. C. & Eaves, C. J. Functional characterization of individual human hematopoietic stem cells cultured at limiting dilution on supportive marrow stromal layers. *Proc. Natl Acad. Sci. USA* **87**, 3584–3588 (1990).
17. Yuan, Y., Shen, H., Franklin, D. S., Scadden, D. T. & Cheng, T. *In vivo* self-renewing divisions of haematopoietic stem cells are increased in the absence of the early G1-phase inhibitor, p18^{INK4C}. *Nature Cell Biol.* **6**, 436–442 (2004).
18. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21^{cip1/waf1}. *Science* **287**, 1804–1808 (2000).
19. Cheng, T., Rodrigues, N., Dombkowski, D., Stier, S. & Scadden, D. T. Stem cell repopulation efficiency but not pool size is governed by p27^{kip1}. *Nature Med.* **6**, 1235–1240 (2000).
20. Orschell-Traycoff, C. M. *et al.* Homing and engraftment potential of Sca-1⁺lin[−] cells fractionated on the basis of adhesion molecule expression and position in cell cycle. *Blood* **96**, 1380–1387 (2000).
21. Wolf, N. S. Dissecting the hematopoietic microenvironment. I. Stem cell lodgment and commitment, and the proliferation and differentiation of erythropoietic descendants in the Sl/Sl^d mouse. *Cell Tissue Kinet.* **7**, 89–98 (1974).
22. Nilsson, S. K. *et al.* Immunofluorescence characterization of key extracellular matrix proteins in murine bone marrow *in situ*. *J. Histochem. Cytochem.* **46**, 371–377 (1998).
23. Kawai, M. *et al.* Development of hemopoietic bone marrow within the ectopic bone induced by bone morphogenetic protein. *Blood Cells* **20**, 191–199 (1994).
24. An, J., Rosen, V., Cox, K., Beauchemin, N. & Sullivan, A. K. Recombinant human bone morphogenetic protein-2 induces a hematopoietic microenvironment in the rat that supports the growth of stem cells. *Exp. Hematol.* **24**, 768–775 (1996).
25. Stier, S., Cheng, T., Dombkowski, D., Carlesso, N. & Scadden, D. T. Notch1 activation increases hematopoietic stem cell self-renewal *in vivo* and favors lymphoid over myeloid lineage outcome. *Blood* **99**, 2369–2378 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Seidman and J. Seidman for access to CaR^{−/−} mice. Financial support for this work was provided by the American Society of Hematology (G.B.A.), the Burroughs Wellcome Fund, the Doris Duke Charitable Trust (D.T.S.) and the National Institutes of Health (M.R.P., E.B. and D.T.S.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.T.S. (scadden.david@mgh.harvard.edu).

LETTERS

Functional genomics reveals genes involved in protein secretion and Golgi organization

Frederic Bard¹, Laetitia Casano¹, Arrate Mallabiabarrena¹, Erin Wallace¹, Kota Saito¹, Hitoshi Kitayama¹, Gianni Guizzunti¹, Yue Hu¹, Franz Wendler², Ramanuj DasGupta³, Norbert Perrimon³ & Vivek Malhotra¹

Yeast genetics and *in vitro* biochemical analysis have identified numerous genes involved in protein secretion^{1,2}. As compared with yeast, however, the metazoan secretory pathway is more complex and many mechanisms that regulate organization of the Golgi apparatus remain poorly characterized. We performed a genome-wide RNA-mediated interference screen in a *Drosophila* cell line to identify genes required for constitutive protein secretion. We then classified the genes on the basis of the effect of their depletion on organization of the Golgi membranes. Here we show that depletion of class A genes redistributes Golgi membranes into the endoplasmic reticulum, depletion of class B genes leads to Golgi fragmentation, depletion of class C genes leads to aggregation of Golgi membranes, and depletion of class D genes causes no obvious change. Of the 20 new gene products characterized so far, several localize to the Golgi membranes and the endoplasmic reticulum.

Drosophila S2 tissue culture cells were transformed to stably express horseradish peroxidase fused to a signal sequence (ss-HRP) on an inducible promoter (Fig. 1a). Addition of Cu²⁺ ions to the cells induced the production of ss-HRP, which is translocated into the endoplasmic reticulum (ER), transported to the Golgi apparatus and then secreted into the medium. An aliquot of the medium was removed to measure peroxidase activity by chemiluminescence, thereby providing a robust assay to monitor secretion in a high-throughput format (Fig. 1b). Two controls were used to verify that secretion of HRP occurred through the generic secretory pathway: knockdown of Syntaxin 5 (a t-SNARE) and β -COP (a component of the COP1 coat)^{3–5} each caused a 100-fold reduction in HRP secretion (Fig. 1c).

We used a genome-wide library of ~22,000 double-stranded RNAs (dsRNAs) that have been previously used in several screens (refs 6–9 and <http://www.flyrnai.org>). Cells secreting ss-HRP were plated in 384-well plates containing dsRNA. After 5 d, ss-HRP production was induced, and 12 h later peroxidase activity released into the supernatant was measured by chemiluminescence (Fig. 2a). Each plate included wells with dsRNA encoding Syntaxin 5 and GFP as positive and negative controls. The screen was carried out in duplicate. Because most dsRNAs did not inhibit HRP secretion, the average for a given plate was very close to that of non-treated wells. Therefore, the z-score of each well, equal to the value of the well (peroxidase activity) minus the average of the plate divided by the standard deviation for the plate, was used to compare the effects of each dsRNA on secretion across the whole set of plates. The average of the two z-scores for each dsRNA is shown in Fig. 2b. The scatter plot of the duplicated assay shows that most dsRNAs did duplicate with a correlation coefficient of 0.63 (Fig. 2c). A few of the dsRNAs did not duplicate, but were

included in the next round of selection to recover potential positives. On the basis of this analysis, 1,133 dsRNAs were selected (Supplementary Table S1).

The genes corresponding to the 1,133 dsRNAs were analysed with Flybase (www.flybase.org). Genes that could affect secretion indirectly through their roles in apoptosis, transcription, protein translation, protein degradation and basic metabolism were discarded from further analysis (Supplementary Table S1). In addition, dsRNAs that scored positively in previous cell survival screens were removed⁶. Known components of the secretory pathway did not score in those cell survival screens⁶. This selection reduced the number to 284 dsRNAs that we tested further in two additional HRP secretion assays in a 96-well plate format. The DNA-binding dye Hoechst was used to exclude dsRNAs that could affect cell number (Fig. 2d). The list of 284 dsRNAs tested and our reasons for excluding 154 from further analysis are given in Supplementary Table S2.

We generated a S2 cell line stably expressing mouse Mannosidase II, a marker of the *cis* and medial Golgi cisternae, coupled to GFP (MannII-GFP) to test the effect of 130 selected genes on Golgi

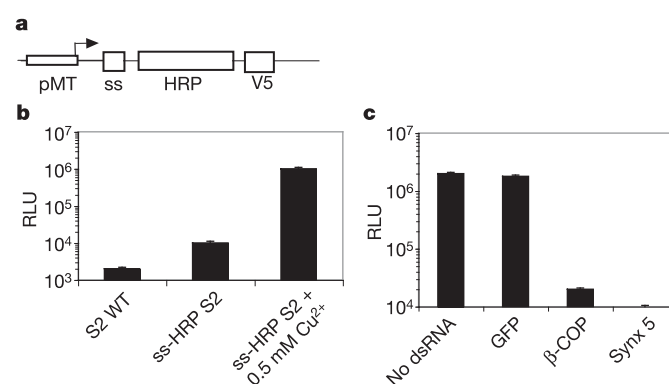


Figure 1 | HRP secretion in *Drosophila* S2 cells. **a**, S2 cells transfected with a plasmid containing the signal sequence (ss) of *Drosophila* Bip appended to HRP and a V5 tag under the influence of an inducible metallothioneine promoter (pMT). **b**, The peroxidase activity in the supernatant from S2 cell culture is HRP. The supernatant from wild-type cells, cells expressing HRP but not induced, and cells induced to produce HRP was analysed for peroxidase activity by chemiluminescence. RLU, relative light units. **c**, RNAi of known effectors of trafficking effectively blocks HRP secretion. RNAi of the *bona fide* transport components β -COP and Syntaxin 5 was used to monitor effects on the secretion of HRP. Cells induced to produce HRP secrete considerable peroxidase activity, which is inhibited on depletion of β -COP and Syntaxin 5. Error bars in **b** and **c** represent the s.d. of triplicate measurements from representative experiments.

¹Cell and Developmental Biology Department, University of California San Diego, La Jolla, California 92093-0634, USA. ²National Institute for Medical Research, the Ridgeway Mill Hill, London NW7 1AA, UK. ³Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

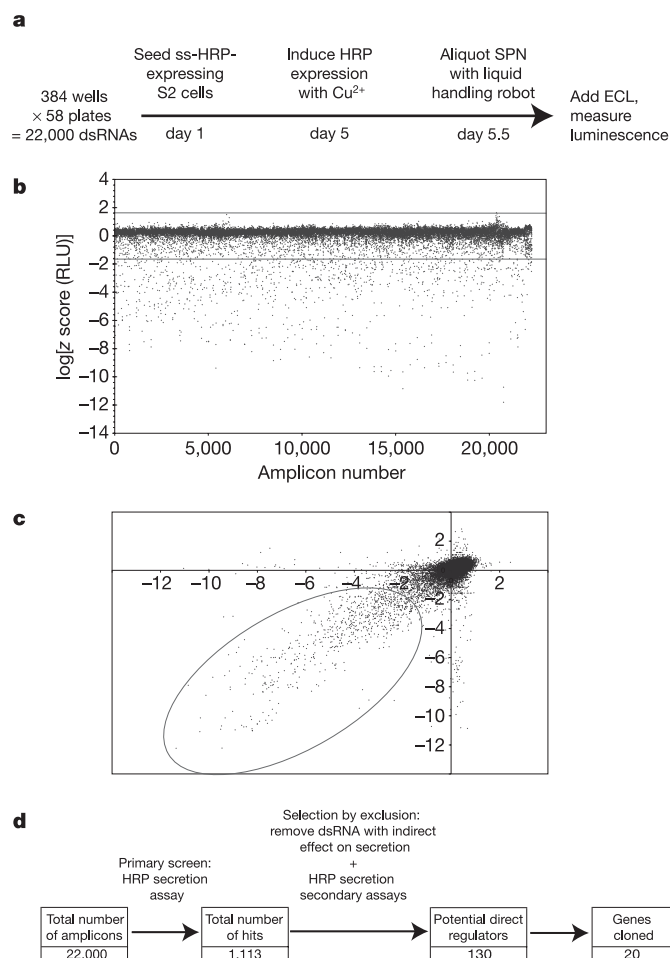


Figure 2 | Identification of the genes involved in HRP secretion. **a**, Each well of each plate containing a dsRNA corresponding to a specific *Drosophila* gene was seeded with 1×10^4 S2 cells stably expressing ss-HRP. After 5 d of incubation, cells were induced to synthesize HRP. After 12 h, 10 μ l of culture supernatant was transferred to another well, the HRP substrate was added, and luminescence was measured. **b**, Whole-genome assay for HRP secretion. The z-score was derived from the luminescence of each well. All dsRNAs that inhibited secretion with a z-score of less than -1.5 (lower red line) were selected as positive hits for further analysis. **c**, Scatter plot for the duplicate screen. The two z-scores derived for each dsRNA are plotted on the x and y axis to show the overall reproducibility. **d**, Flow chart showing the selection of genes involved in secretion. From an initial collection of 1,133 genes (Supplementary Table S1), 130 (Supplementary Table S3) were found to be highly specific for HRP secretion. The subcellular localization of 20 of these genes was assessed after cloning and transient transfection.

organization. As previously reported¹⁰, Golgi membranes in S2 cells are organized as several unconnected stacks of cisternae (Fig. 3a). We incubated S2 cells expressing MannII-GFP with dsRNAs by the procedure described above for HRP-secreting S2 cells, and then imaged them by high-resolution deconvolution fluorescence microscopy. The genes were classified into four groups on the basis of the effect of their depletion on Golgi membranes in greater than 50% of the total cells. RNA-mediated interference (RNAi) of class A genes fused Golgi membranes with the ER, as shown by the relocation of MannII-GFP in a ring around the nucleus and a diffuse reticular network (Fig. 3b). RNAi of class B genes fragmented the Golgi membranes into smaller elements, RNAi of class C genes caused aggregation and swelling, and RNAi of class D genes had no apparent effect on Golgi organization (Fig. 3b). The complete list of the four classes of genes, their potential human orthologues and their domains are given in Supplementary Table S3. We found that 77 out of the 130 genes had potential human orthologues and 26 had homologues identified previously as trafficking components. The dsRNAs that have potential off-target effects on the basis of the presence of a 21-base-pair overlap with other genes, and which therefore need further validation, are also listed in Supplementary Table S3.

In mammalian cells, the Golgi membranes are fragmented and protein transport is blocked during mitosis¹¹. The list of 130 genes includes *separase*, *k1p61F* and *microtubule star*, which have been linked to mitosis-specific events^{12–14}. It is therefore possible that effects on trafficking and Golgi organization in S2 cells depleted of these genes are due to arrest in mitosis. RNAi-treated MannII-GFP cells were visualized with antibody against tubulin and mitosis-specific antibody against phosphorylated histone H3. We could detect mitotic cells in control wells (Supplementary Fig. S1), but there was no increase in the mitotic index of cells depleted of the genes mentioned above. Consistently, the phenotypes of class B or C genes were not associated with DNA condensation, microtubule reorganization or increase in staining for phosphorylated histone H3 (Supplementary Fig. S1). Therefore, the fragmented Golgi phenotype and inhibition in secretion of HRP in these cells are not due to an arrest in mitosis. However, other caveats remain; for example, we found that the knockdown of genes involved in fatty acid and cholesterol biosynthesis results in a block of HRP secretion and a class A Golgi membrane phenotype (Supplementary Table S2). To confirm their direct role in membrane trafficking, therefore, the 104 candidate genes need further functional tests and, notably, characterization of the intracellular localization of their products.

To this end, 20 randomly selected genes including *CG14181*, a likely orthologue of *Use1* (encoding a t-SNARE localized to the ER in *Saccharomyces cerevisiae*¹⁵), were cloned in an inducible expression vector with a V5 tag. S2 cells expressing MannII-GFP were transfected with the tagged cloned genes, and the gene products were

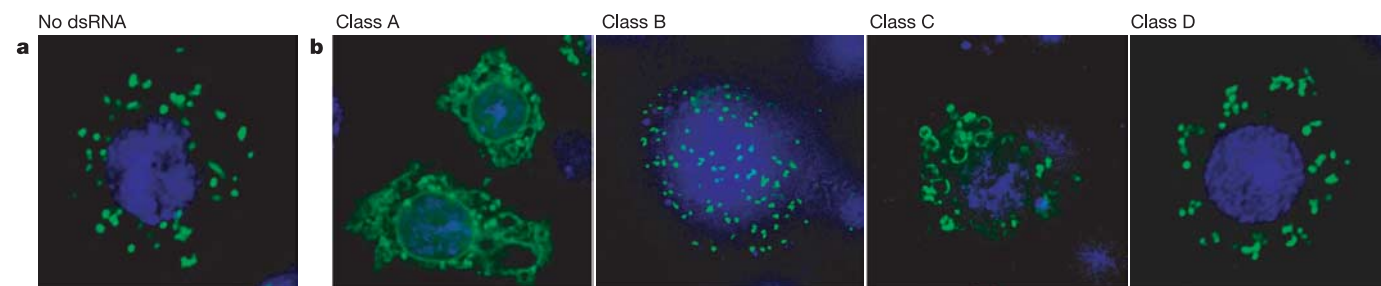


Figure 3 | *Drosophila* genes involved in Golgi organization. S2 cells stably expressing MannII-GFP, a marker of medial Golgi, were subjected to RNAi for each of the 130 genes identified to be essential for secretion. **a**, Organization of the untreated control MannII-GFP in S2 cells. MannII appears in discrete units around the nuclear periphery. **b**, On the basis of

their effects on organization of the Golgi membranes, the 130 secretory components are grouped into four classes (Supplementary Table S3). Their depletion caused the Golgi to fuse with the ER (class A), fragmented the Golgi (class B), induced Golgi membranes to aggregate and swell (class C) or had no effect on Golgi organization (class D).

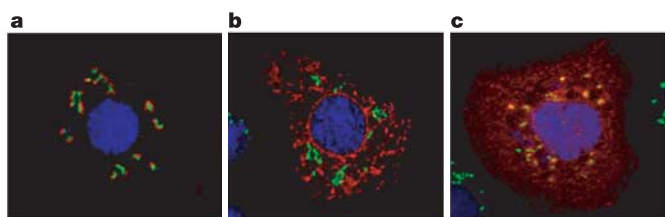


Figure 4 | Localization of the products of new genes regulating secretion. Candidate genes were cloned in an inducible vector with a V5 tag and transiently transfected into S2 cells expressing MannII–GFP. After fixation, cells were labelled with Hoechst (to stain DNA) and an anti-V5 antibody coupled to Texas red. MannII–GFP is green, DNA is blue and the gene product is red. **a**, CG11098 (Golgi). **b**, CG32675 (ER). **c**, CG8309 (Golgi and cytosol).

visualized with an antibody against V5 and imaged by deconvolution. The localization of the gene products was compared with that of the Golgi marker (MannII–GFP), the ER pattern, and the diffuse cytosolic pattern of a soluble protein (examples of typical localization are shown in Fig. 4). Of the 20 cloned genes products, 4 localized to the Golgi membranes and 7 (including the CG14181 product) localized to the ER, suggesting that they have a direct role in membrane trafficking (Table 1). The gene product of CG11098, named *TANGO1* for transport and Golgi organization (Table 1), is perfectly juxtaposed to the MannII–GFP-containing Golgi membranes (Fig. 2a). *TANGO1* contains an amino-terminal Src homology 3 (SH3) domain followed by two coiled-coil domains, two transmembrane domains and a proline-rich domain. The coiled-coil domains have homology to the yeast protein Uso1p and the golgin p115. *TANGO1*, however, is not the *Drosophila* orthologue (encoded by CG1422) of Uso1p or p115 and has a human but not an *S. cerevisiae* homologue.

Another example of a metazoan-specific gene is CG1098, whose product localizes to the cytoplasm and Golgi membranes. CG1098 has a human orthologue, *NRBP*, which encodes a serine/threonine kinase that is recruited to Golgi membranes on viral infection¹⁶ and that, when overexpressed, perturbs early Golgi membranes in mammalian cells¹⁷. Among the candidates with a yeast homologue, *TANGO2* (CG11176) contains a domain, DUF833, of unknown

function that is widely conserved, even in some bacteria. *TANGO4* (CG1796) also has a potential homologue in yeast, the splicing factor gene *Prp46*. Contrary to other splicing factors, however, RNAi of *TANGO4* resulted in a marked effect on Golgi membranes (class A phenotype) and its gene product localized to the cytosol and not the nucleus, suggesting that *TANGO4* regulates secretion independently from its potential role in RNA splicing. The challenge now is to understand how these and the other identified components regulate protein secretion and whether they regulate Golgi membrane organization directly. Their characterization will hopefully help to resolve the emerging complexities of the metazoan secretory pathway.

METHODS

Constructs and cell lines. The *HRP-C* gene was cloned by PCR from a construct¹⁸ provided by D. Cutler (MRC, University College London) and inserted into pMT/BiP/V5–His (Invitrogen). The sequence corresponding to the 100 N-terminal amino acids of mouse MannII was fused with the GFP sequence and inserted into pAC5/V5–His. S2 cells were cotransfected with pHygro (Invitrogen) in a ratio of 20 to 1 and selected with 0.3 mg ml^{−1} of Hygromycin.

Primary screen and analysis. Two sets of 58 plates containing 0.25 µg of dsRNA per well were provided by the *Drosophila* RNAi Screening Center (DRSC; <http://flyrnai.org>). 1×10^4 ss-HRP cells were seeded in each well with a Multidrop 384. After incubation for 2 h, 20 µl of fetal bovine serum containing medium was added. After 5 d, the culture medium was replaced with 50 µl of medium containing 500 µM copper and the cells were incubated overnight. We transferred 10 µl of supernatant into a receptacle plate with a Cybio CyBi-Well vario system and 50 µl of ECL reagent (Perkin-Elmer Western Lightning) was added. Luminescence was measured with an Analyst plate reader. Each receptacle plate was assigned a number and a barcode similar to the initial set for automatic identification by the plate reader.

Analysis of primary screen. z-Scores were derived from the log value of luminescence and genes with a score below −1.5 were selected. By using the DSRC database, genes that scored in cell survival screens⁶, as well as genes involved in transcription, RNA splicing, protein translation and proteasome function, were excluded from further analysis. The raw data for the whole screen will be made available (<http://flyrnai.org>). Identification of potential orthologues in humans was based on information available in Flybase (<http://flybase.bio.indiana.edu/>) and on the reciprocal best blast searches with the InParanoid algorithm.

Secondary screens and morphological effects on Golgi membranes. Using PCR templates provided by the DRSC, we resynthesized dsRNAs and tested their effect on HRP secretion with a protocol similar to the primary screen. To measure cell number, Hoescht was added at 5 µg ml^{−1} to cells for 90 min, the cells were washed, and ultraviolet fluorescence was measured with a plate reader (Tecan). dsRNAs resulting in greater than 50% inhibition of HRP secretion without affecting cell number were selected as positives. MannII–GFP cells were incubated with dsRNAs as described above; after 5 d, the cells were transferred to concanavalin-treated 96-well glass-bottom plates, allowed to spread for 2 h and then fixed and labelled with Hoescht at 2 µg ml^{−1}. Image stacks for five fields in each well were acquired with a 60× objective and treated for deconvolution.

Cloning of genes for expression in S2 cells. Genes were cloned by RT–PCR from a library of poly(A) RNA from *Drosophila* larvae (Clontech) using the Gateway system (Invitrogen) and subcloned into pDEST48. MannII–GFP S2 cells were transfected 2 d before gene expression was induced with Cu²⁺ for 4 h, fixed, labelled with antibody against V5 (Invitrogen), and processed for imaging as described above.

Received 22 August; accepted 18 October 2005.

- Novick, P. & Schekman, R. Secretion and cell-surface growth are blocked in a temperature-sensitive mutant of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **76**, 1858–1862 (1979).
- Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
- Nichols, B. J. & Pelham, H. R. SNAREs and membrane fusion in the Golgi apparatus. *Biochim. Biophys. Acta* **1404**, 9–31 (1998).
- Duden, R., Allan, V. & Kreis, T. Involvement of β-COP in membrane traffic through the Golgi complex. *Trends Cell Biol.* **1**, 14–19 (1991).
- Serafini, T. et al. A coat subunit of Golgi-derived non-clathrin-coated vesicles with homology to the clathrin-coated vesicle coat protein β-adaptin. *Nature* **349**, 215–220 (1991).
- Boutros, M. et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).

Table 1 | Intracellular localization of the gene products involved in HRP secretion

CG number	Gene name	Golgi phenotype	Localization of gene product
CG11098	<i>TANGO1</i>	A	Golgi
CG11176	<i>TANGO2</i>	A	cyto + Golgi
CG12444	<i>TANGO3</i>	A	ER
CG14181	<i>Use1</i>	A	ER
CG1796	<i>TANGO4</i>	A	cyto
CG32675	<i>TANGO5</i>	A	ER
CG18398	<i>TANGO6</i>	B	cyto + Golgi
CG8309	<i>TANGO7</i>	B	cyto + Golgi
CG14503	<i>TANGO8</i>	C	ER
CG9191	<i>Klp61F</i>	C	ER + Golgi + microtubules
CG10007	<i>TANGO9</i>	D	Golgi
CG1098	<i>Madm</i>	D	cyto + Golgi
CG1841	<i>TANGO10</i>	D	cyto
CG30404	<i>TANGO11</i>	D	ER
CG31052	<i>TANGO12</i>	D	Golgi
CG32632	<i>TANGO13</i>	D	Golgi
CG33553	<i>Doa</i>	D	ER
CG4775	<i>TANGO14</i>	D	ER
CG7850	<i>puckered</i>	D	ER + Golgi
CG8588	<i>pastrel</i>	D	cyto

Twenty genes were cloned and tagged with V5, and their subcellular localization was monitored in S2 cells expressing MannII–GFP. The localization of the gene product is abbreviated as follows: cyto, cytosolic; Golgi, Golgi membranes; ER, endoplasmic reticulum. Computed genes without a previous name are labelled *TANGO* for transport and Golgi organization.

7. DasGupta, R., Kaykas, A., Moon, R. T. & Perrimon, N. Functional genomic analysis of the Wnt-wingless signalling pathway. *Science* **308**, 826–833 (2005).
8. Eggert, U. S. *et al.* Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol.* **2**, e379 (2004).
9. Kiger, A. A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27 (2003).
10. Stanley, H., Botas, J. & Malhotra, V. The mechanism of Golgi segregation during mitosis is cell type-specific. *Proc. Natl Acad. Sci. USA* **94**, 14467–14470 (1997).
11. Shorter, J. & Warren, G. Golgi architecture and inheritance. *Annu. Rev. Cell. Dev. Biol.* **18**, 379–420 (2002).
12. Uhlmann, F. Separase regulation during mitosis. *Biochem. Soc. Symp.* **70**, 243–251 (2003).
13. Sharp, D. J. Cell division: MAST sails through mitosis. *Curr. Biol.* **12**, R585–R587 (2002).
14. Snaitch, H. A., Armstrong, C. G., Guo, Y., Kaiser, K. & Cohen, P. T. Deficiency of protein phosphatase 2A uncouples the nuclear and centrosome cycles and prevents attachment of microtubules to the kinetochore in *Drosophila microtubule star (mts)* embryos. *J. Cell Sci.* **109**, 3001–3012 (1996).
15. Dilcher, M. *et al.* Use1p is a yeast SNARE protein required for retrograde traffic to the ER. *EMBO J.* **22**, 3664–3674 (2003).
16. Chua, J. J., Ng, M. M. & Chow, V. T. The non-structural 3 (NS3) protein of dengue virus type 2 interacts with human nuclear receptor binding protein and is associated with alterations in membrane structure. *Virus Res.* **102**, 151–163 (2004).
17. De Langhe, S., Haataja, L., Senadheera, D., Groffen, J. & Heisterkamp, N. Interaction of the small GTPase Rac3 with NRBP, a protein with a kinase-homology domain. *Int. J. Mol. Med.* **9**, 451–459 (2002).
18. Connolly, C. N., Futter, C. E., Gibson, A., Hopkins, C. R. & Cutler, D. F. Transport into and out of the Golgi complex studied by transfecting cells with cDNAs encoding horseradish peroxidase. *J. Cell Biol.* **127**, 641–652 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Malhotra laboratory for discussions; members of the DRSC for advice; the Institute for Chemistry and Cell Biology for use of their Cybio robot; and J. Feramisco and members of the UCSD Cancer Center imaging facility for help with microscopy. Work in the Malhotra laboratory is supported by NIH grants and a senior investigator award from Sandler's Program for Asthma Research. N.P. is a Howard Hughes investigator.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.M. (malhotra@biomail.ucsd.edu).

LETTERS

Gene network shaping of inherent noise spectra

D. W. Austin^{1,3}, M. S. Allen^{1,3}, J. M. McCollum³, R. D. Dar¹, J. R. Wilgus³, G. S. Saylor³, N. F. Samatova², C. D. Cox³ & M. L. Simpson^{1,3}

Recent work demonstrates that stochastic fluctuations in molecular populations have consequences for gene regulation^{1–10}. Previous experiments focused on noise sources or noise propagation through gene networks by measuring noise magnitudes. However, in theoretical analysis, we showed that noise frequency content is determined by the underlying gene circuits, leading to a mapping between gene circuit structure and the noise frequency range^{11,12}. An intriguing prediction from our previous studies was that negative autoregulation shifts noise to higher frequencies where it is more easily filtered out by gene networks¹¹—a property that may contribute to the prevalence of autoregulation motifs (for example, found in the regulation of ~40% of *Escherichia coli* genes). Here we measure noise frequency content in growing cultures of *E. coli*, and verify the link between gene circuit structure and noise spectra by demonstrating the negative autoregulation-mediated spectral shift. We further demonstrate that noise spectral measurements provide mechanistic insights into gene regulation, as perturbations of gene circuit parameters are discernible in the measured noise frequency ranges. These results suggest that noise spectral measurements could facilitate the discovery of novel regulatory relationships.

We investigated single gene circuits on high copy number plasmids (pGFPasv) in *E. coli* TOP10 cells where destabilized (half-life ~110 min) green fluorescent protein (GFP) was constitutively expressed (Fig. 1a). The average GFP fluorescence, which corresponded to the concentration of mature GFP protein, was measured in individual cells in growing cultures for 4–8 h periods using time-lapse microscopy^{1,2} (see the Supplementary Movie). Cells remained in the exponential growth phase throughout the experiment. Fluorescence was recorded every 5 min through multiple generations of cell division (Fig. 1b), and noise was found as the difference between the fluorescence of individual cells and the population mean determined at each measurement time (Fig. 1c). Individual noise traces (trajectories) that spanned the entire growth time were constructed by combining sequential noise traces of cells through lines of descent (Fig. 1b and Supplementary Information). Normalized autocorrelation functions of noise in fluorescence were estimated for individual trajectories ($\Phi_m(\tau)$) and composites ($\Phi_c(\tau)$) of all tracked trajectories in each cell culture for the duration of the experiment (Fig. 1d). The composite autocorrelation functions provided the better estimate of the underlying random process, while individual trajectory autocorrelation functions may provide insights into gene circuit structure or function as described below.

Histograms of noise frequency ranges (see Box 1) extracted from the individual trajectory autocorrelation functions (Fig. 2a, b) were compiled and compared with noise frequency range distributions found from exact stochastic simulation^{13,14} using a model that included intrinsic and extrinsic noise, protein dilution and protein decay (see the Supplementary Information). The simulations produced as much data as 500 separate experiments, and the resulting

distributions estimated the probability of finding a given noise frequency range from a randomly selected trajectory. Experimental distributions were drawn from this random process and were subject to variation, particularly the relatively rare events in the high-frequency tails. Although some measured distributions suggested a bimodal distribution (Fig. 2), this was probably due to non-representative sampling of the rare high-frequency events (see the Supplementary Information).

Analytical models predict that protein dilution and decay rates are dominant factors defining the noise frequency range in constitutively expressed gene circuits¹¹ (Box 1). To determine noise frequency range sensitivity to protein dilution, we varied cell growth rate for the pGFPasv circuits by controlling temperature, and in a separate experiment we changed the protein decay rate using a plasmid (pGFPaav) containing a reduced half-life (~60 min) GFP variant¹⁵. These perturbations to gene circuit parameters were clearly visible in the noise spectral measurements, as noise frequency ranges extended to higher frequencies as a result of faster growth (Fig. 2a) or higher

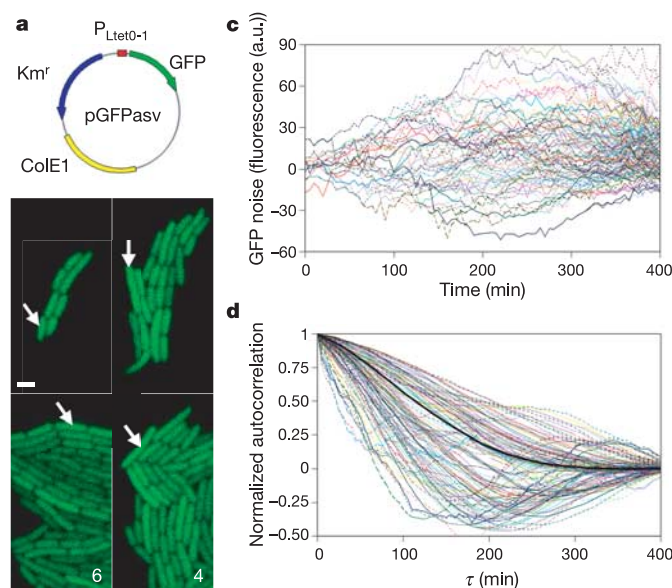


Figure 1 | Measurement of noise frequency ranges using fluorescence time-lapse microscopy. **a**, Plasmid pGFPasv containing the constitutively expressed GFP (110-min half-life) gene circuit. **b**, Bi-hourly (time in hours shown at the bottom right of images) snapshots of fluorescence in pGFPasv cell cultures (doubling time = 59 min; scale bar = 2 μ m). The arrows follow one trajectory through six cell divisions. **c**, **d**, Noise in GFP concentration (**c**) and normalized autocorrelation functions (**d**) for all trajectories tracked from the experiment shown in **b**. The composite autocorrelation is shown as a bold black curve. Fluorescence in **c** is given in arbitrary units (a.u.).

¹Molecular-Scale Engineering and Nanoscale Technologies Group, and ²Computer Science and Mathematics Division, Oak Ridge National Laboratory, Bethel Valley Road, Oak Ridge, Tennessee 37831, USA. ³University of Tennessee, Knoxville, Tennessee 37996, USA.

Box 1 | Noise frequency range

The noise frequency range is determined by the nature of the noise sources and gene circuit filtering. Fluctuations generated within the gene circuit (intrinsic noise) arise from the random timing and discrete nature of molecular transitions (for example, transcription, translation, decay or dilution through growth, and multimerization^{9,10}). Extrinsic noise arises from the intrinsic noise of global sources (for example, fluctuations in ribosome or RNA polymerase populations) or of upstream genes in a regulatory cascade⁷. Intrinsic noise sources have wide frequency ranges, while extrinsic sources are confined to a lower-frequency range by their own gene circuit filtering².

The frequency range is found from the normalized autocorrelation function ($\Phi(\tau)$) of noise in a reporter protein concentration. Total noise in the protein concentration is produced by the sum of intrinsic and extrinsic noise source components filtered primarily by protein dilution through cell growth (rate = δ) and protein decay (rate = γ) such that

$$\Phi(\tau) = W_E \left(\frac{\delta + \gamma}{\gamma} e^{-\delta\tau} - \frac{\delta}{\gamma} e^{-(\delta + \gamma)\tau} \right) + W_I e^{-(\delta + \gamma)\tau}$$

where W_E and W_I account for the relative contributions of extrinsic and intrinsic noise sources, and we have assumed that the frequency range of the extrinsic noise source is set by the cell growth rate².

We define the noise frequency range (F_N) as

$$F_N = \frac{1}{\tau_{1/2}}$$

where $\tau_{1/2}$ is the value of τ where $\Phi(\tau)$ drops to $\frac{1}{2}$. Slower fluctuations remain correlated over longer periods and therefore have lower values of F_N .

Theoretical analysis predicts that negative autoregulation increases the noise frequency range such that¹¹:

$$F_{N,\text{autoreg}} = (1 + |T|)F_{N,\text{unreg}}$$

T is a unitless parameter proportional to the degree to which the autoregulated gene circuit resists changes in the equilibrium protein concentration (strength of regulation)¹¹. The measured noise frequency range can be used to determine the strength of regulation using the equation above.

protein decay rate (Fig. 2b). Varying temperature changes the rates of all reactions, and while noise magnitude is sensitive to all of these changes, the noise frequency range of constitutively expressed circuits is mainly sensitive to protein dilution and decay rates only (see the Supplementary Information).

Significant variation from the noise frequency range predicted from decay and dilution rates indicates additional structure beyond the simple components of the constitutively expressed gene circuit. Most of the effects that have been considered (for example, extrinsic noise^{1,2}, protein–DNA binding in transcriptional control^{12,16}, protein dimerization¹⁷, GFP maturation (see the Supplementary Information)) lower the noise frequency range. To test the predicted increased noise frequency range with negative autoregulation¹¹, we constructed circuits with the gene for the protein TetR inserted upstream of *gfp*, creating a transcriptional fusion (pTetR–GFPasv; Fig. 3a). This circuit was negatively autoregulated, as its expression was repressed by TetR binding to operator sites in the promoter¹⁸. A control circuit without autoregulation was also tested, in which a chromosomal copy of *tetR* was constitutively expressed from the P_{N25} promoter. In both cases, repression was relieved by addition of anhydrotetracycline (ATc) to the growth medium, and thus allowed the modulation of GFP expression (Fig. 3b).

To determine if ATc had an effect on noise spectra independent of the autoregulation of the TetR circuit, we measured the noise frequency range of the pGFPasv circuits in media supplemented with 100 ng ml⁻¹ of ATc. There was a marked modification in the noise frequency range distribution (Fig. 3c), indicating a change in either the processing of the noise or the nature of the noise sources.

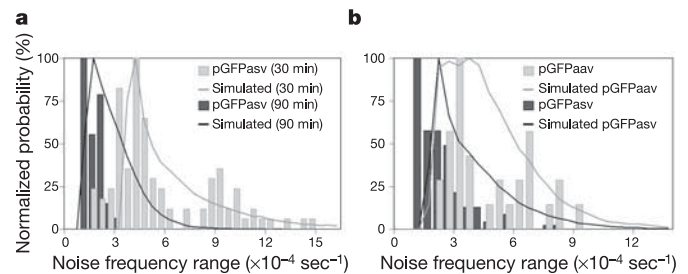


Figure 2 | Effects of cell doubling time and protein half-life on noise frequency range. Measured distributions are shown as vertical bars and simulated distributions as solid lines. **a**, Shift in noise frequency range for the pGFPasv circuit as doubling time increases from ~30 min (100 trajectories; $T = 32^\circ\text{C}$) to ~90 min (120 trajectories; $T = 22^\circ\text{C}$). **b**, Shift in noise frequency range as protein decay time decreases from 110 min (pGFPasv; 59-min doubling time; 154 trajectories; $T = 26^\circ\text{C}$) to 60 min (pGFPaav; 56-min doubling time; 33 trajectories; $T = 26^\circ\text{C}$).

Our modelling points to the latter, with ATc inhibition of translation leading to a whitening of extrinsic noise, which we then explored using a stochastic simulation model that included ribosome–ATc heterodimer formation (see the Supplementary Information). The frequency range distribution extracted from these simulations was compared to the measured distribution (Fig. 3c), with both showing a characteristic peak shift and peak broadening. Although not conclusive, this gross agreement between measured and simulated distributions supports the hypothesis that the mechanism of ATc-mediated noise frequency range modulation is an increase in high-frequency content of the global extrinsic noise associated with translation, and a reduction of the weighting of extrinsic noise.

We measured noise frequency range distributions of pTetR–GFPasv and the control cells grown in media with 100 ng ml⁻¹ of ATc. Composite noise frequency ranges of the negatively autoregulated pTetR–GFPasv exceeded those of the constitutively expressed pGFPasv in 100 ng ml⁻¹ of ATc by as much as ~2–3-fold (Figs 3d and 4a), whereas the control circuits showed no noise frequency range increase (Fig. 4a). The negative autoregulation-mediated noise remodelling was seen as an increase of the noise frequency range (Fig. 4a) and as a modification of the shape of the distribution (Fig. 3d and Supplementary Information). Autoregulation frequency response is limited by protein decay and dilution, and therefore has a larger effect on slower fluctuations than on faster fluctuations. Noise trajectories that would have clustered at the lower end of the frequency range distribution in unregulated cells are pushed to higher values by negative autoregulation, while those in the higher-frequency tail of the distribution are only weakly affected (Fig. 3e). This results in frequency range distributions closer to normal distributions (Fig. 3d). The frequency shift and the change in distribution shape are indicative of the presence of negative autoregulation.

The magnitude of the autoregulation-mediated frequency shift was indicative of the strength of regulation (Box 1), which varied significantly as a function of cell doubling time (Fig. 4a). The measurements indicate that regulation strength is small at both fast and slow rates of cell growth, with a peak at intermediate levels of cell growth. In our gene circuit model (Fig. 4b), regulation strength is the product of terms that describe (1) the change in free (not bound to ATc and capable of repression) TetR dimer concentration in response to changes in transcription rate, and (2) the change of transcription rate in response to changes in free TetR dimer binding to the operator (see the Supplementary Information). Our measurements consistently showed lower average fluorescence at high cell growth rates (data not shown), indicating that GFP (and probably TetR) concentration was reduced by rapid dilution. At high growth

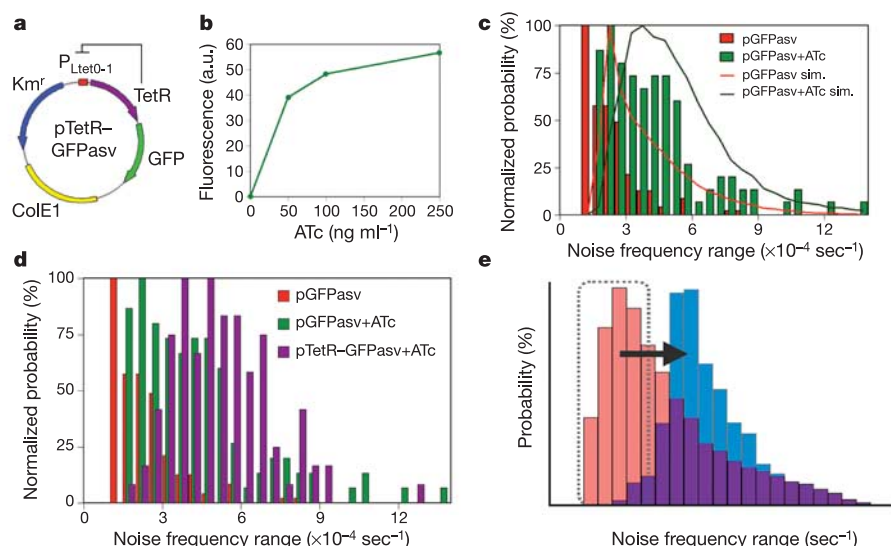


Figure 3 | Effect of negative autoregulation on noise frequency range. **a**, pTetR-GFPasv negatively autoregulated gene circuit with **b**, repression strength modulated by ATc. **c**, Effect of ATc on the noise frequency range of the unregulated pGFPasv circuit (doubling time ~ 60 min; 154 trajectories without ATc; 114 trajectories with ATc). **sim.**, simulated. **d**, Negative autoregulation-mediated shift of noise frequency range (doubling time ~ 60 min; pGFPasv: 154 trajectories without ATc, 114 trajectories with ATc; pTetR-GFPasv: 114 trajectories). **e**, Model of the shift of frequency range distribution shape due to negative feedback. The red bars represent an unregulated circuit distribution; blue bars represent distribution for the circuit with negative autoregulation. The dashed box and arrow show the shift of the low-frequency trajectories to the centre of the distribution while the higher-frequency trajectories are unaffected. Fluorescence in **b** is given in arbitrary units (a.u.).

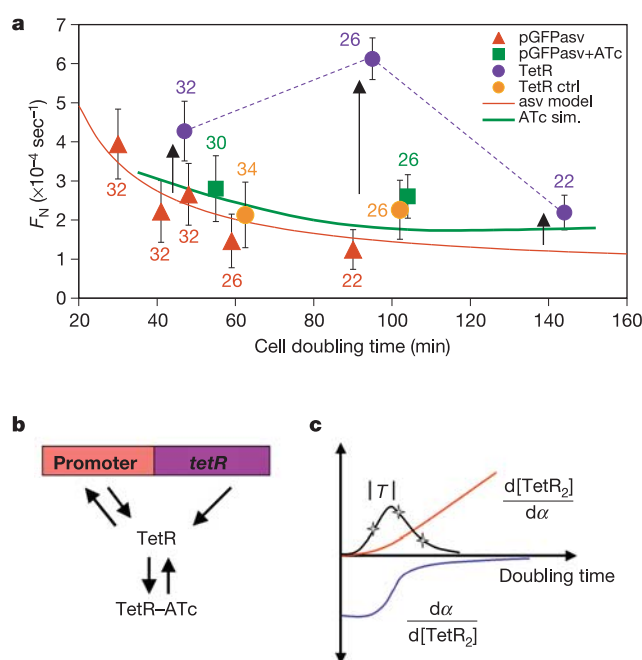


Figure 4 | Regulation strength modulation of noise frequency range. **a**, Noise frequency range versus doubling time. Measured points are shown with $\pm 1\sigma$ error bars estimated from simulation (see the Supplementary Information). The red line is the analytical curve (see Box 1 and the Supplementary Information) for the pGFPasv circuit, and was found from the analytical expression for the autocorrelation function (see Box 1). The green line is the simulated curve (see the Supplementary Information) for pGFPasv + 100 ng ml^{-1} ATc and was found from the simulation of the pGFPasv circuit with ATc-ribosome binding (see the Supplementary Information). Vertical black arrows represent regulation strength determined by the shift of the noise frequency range. The temperature (in $^{\circ}\text{C}$) of each experiment is indicated by each data point. The TetR data points are for the circuit with autoregulated *tetR* expression, while the TetR ctrl data points are for the circuit with constitutive *tetR* expression. **b**, **c**, Regulation of the pTetR-GFPasv circuit. The red curve shows the concentration of free TetR dimer ($[\text{TetR}_2]$) variation with transcription rate (α); the blue curve shows transcription rate variation with $[\text{TetR}_2]$; and the black curve shows net regulation strength. The stars illustrate points on the regulation curve similar to the TetR data in **a**.

rates, repression strength was low as most TetR was bound with ATc. Conversely, at low cell growth rates, abundant free TetR dimer was formed, but regulation strength was low as the repression curve saturated^{12,16}. The result was a regulation curve that peaked at intermediate cell growth rates where there was an adequate abundance of free TetR dimer, but not so much that the repression curve had saturated (Fig. 4c).

Previous work demonstrated the noise magnitude damping effect of negative autoregulation¹⁹. We have demonstrated that gene circuits also manipulate noise spectra, impacting the regulatory effect of the noise as it propagates through the gene network. We verified the link between gene circuit structure and the noise frequency range, and showed that changes in gene circuit parameters (for example, cell growth and protein decay rates) modulate the noise frequency range. Our results show a shift of noise spectra to higher frequencies and a remodelling of the noise frequency range distribution that is characteristic of negative autoregulation. This frequency shift may have biological relevance, as higher-frequency noise is more easily filtered out by downstream gene circuits in a regulatory cascade, and therefore has little regulatory impact¹¹. Noise may play a beneficial role in some gene circuit functions^{17,20,21}, and circuit structure may have evolved to optimize rather than minimize noise. The theory described here predicts that positive regulation, which plays an important role in many gene circuit functions^{17,22,23}, increases noise magnitude while shifting noise frequency range downward into a more regulatory-relevant regime where it may play a role in the function of some genetic switching elements¹⁷. Intriguingly, the noise frequency range distributions provided a means for evaluating a hypothesis of the mechanism of ATc-mediated remodelling of noise spectra in unregulated gene circuits (see the Supplementary Information), suggesting that the coupling of spectral measurements and simulation provides a new tool for probing mechanistic details of gene circuit regulation.

METHODS

Synthetic gene circuits. The high-copy plasmid pGFPasv (pZE21-GFPasv; ref. 24) was used both directly and as the platform for further genetic manipulation. This plasmid contains the *gfpmut3** reporter gene, modified to include a carboxy-terminal amino-acid tag previously shown to facilitate accelerated degradation of GFP protein within the cell¹⁵, and is expressed from the strong promoter $P_{\text{Ltet0-1}}$ (ref. 18). GFP variants with alternative half-lives were created by replacement of the 3' tag of the *gfp* gene with synthetic oligonucleotides using standard methods¹⁵. The *tetR* gene was amplified by polymerase chain reaction (PCR), cloned, sequenced and then subsequently digested and inserted upstream of *gfp-asv* to form the ATc-inducible, negatively autoregulated transcriptional fusion in plasmid pTetR-GFPasv. The

non-autoregulated control system was created by transforming the pGFPasv plasmid into *E. coli* DH5 α PRO (BD Biosciences), which contains a chromosomal insertion of the *tetR* gene behind the P_{N25} promoter.

Cultures, growth conditions and imaging. *E. coli* TOP10 (Invitrogen) cells were used for all experiments unless otherwise noted. For slide experiments, overnight cultures were grown at 28 °C in M9 medium supplemented with 100 mM leucine and 10% (v/v) Luria–Bertani broth, plus kanamycin (50 μ g ml⁻¹) and ATc (anhydrotetracycline; Acros) when appropriate. Cultures were transferred 1:3 into fresh media and allowed to recover for ~1 h before spreading 10 μ l onto freshly prepared glass microscope slides coated with the same media containing 1% (w/v) Sea Plaque low-melt agarose (FMC). Coverslips were applied and the slides were allowed to incubate 1 h at 28 °C before imaging. Slides were imaged every 5 min using the Leica SP2 confocal system with an excitation wavelength of 488 nm and emission wavelength at 500–564 nm. Temperature was maintained using a heating lamp and monitored with a Hanna KJF thermocouple.

Data processing. Noise traces were extracted from the images for nearly all possible trajectories in each experiment. Custom Matlab (Mathworks) programs were used to find mean fluorescence of an entire cell population and to estimate population doubling time from an exponential growth curve. Normalized autocorrelations functions (ACFs) for individual trajectories (Φ_m) were found from the noise time series ($X_m(nT_s)$) using a biased algorithm²⁵

$$\Phi_m(jT_s) = \frac{\sum_{n=1}^{N-j} X_m(nT_s)X_m((n+j)T_s)}{\sum_{n=1}^N X_m^2(nT_s)}$$

where T_s was the five-minute sampling interval, n was the sample number (1, 2, ..., n), and j had integer values from 0 to $n - 1$. The composite autocorrelation function (Φ_c) for M cell trajectories was found using:

$$\Phi_c(jT_s) = \frac{\sum_{m=1}^M \sum_{n=1}^{N-j} X_m(nT_s)X_m((n+j)T_s)}{\sum_{m=1}^M \sum_{n=1}^N X_m^2(nT_s)}$$

Simulations. Gillespie's stochastic simulation algorithm¹³ was used to generate simulated estimates of ACFs, error bars, F_N histograms, and F_N behaviour as a function of doubling time. The models accounted for extrinsic and intrinsic noise sources, binding of 30S ribosomal subunits by ATc, appropriate temperature corrections, and protein sinks consisting of decay and dilution via stochastically timed cell division processes (see the Supplementary Information). Individual simulations were designed to closely match experimental conditions in terms of numbers of cells tracked, distribution of cell division frequencies, and experimental duration. Sets of many individual simulations were used to define mean system behaviour, its variability, and experimental uncertainty.

Received 23 April; accepted 5 September 2005.

- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
- Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
- Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
- Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
- Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
- Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
- Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: From theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
- Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).
- Simpson, M. L., Cox, C. D. & Saylor, G. S. Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl Acad. Sci. USA* **100**, 4551–4556 (2003).
- Simpson, M. L., Cox, C. D. & Saylor, G. S. Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *J. Theor. Biol.* **229**, 383–394 (2004).
- Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
- Gibson, M. A. & Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104**, 1876–1889 (2000).
- Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
- Kepler, T. B. & Elston, T. C. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
- Cox, C. D. *et al.* Analysis of noise in quorum sensing. *OMICS* **7**, 317–334 (2003).
- Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I₁-I₂ regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
- Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
- Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
- Thattai, M. & van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523–530 (2004).
- Becskei, A., Seraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
- Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100**, 7714–7719 (2003).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
- Bendat, J. S. & Piersol, A. G. *Data: Analysis and Measurement Procedures* (Wiley, New York, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Golding and J. Dunlap for advice on sample preparation and imaging, and M. Elowitz, J. Collins and T. Gardner for the gift of plasmids. This work was supported by the National Academies Keck Futures Initiative, the DARPA Bio-Computation Program, the NSF, the DOE Office of Advanced Scientific Computing Research, and was a user project of the Oak Ridge National Laboratory (ORNL) Center for Nanophase Materials Sciences (CNMS). D.W.A. acknowledges support from an ORNL CNMS Research Scholar Fellowship. R.D.D. acknowledges support from the DOE Science Undergraduate Laboratory Internship program.

Author Contributions D.W.A., M.S.A., J.M.M., C.D.C. and M.L.S. planned the experimental, analytical and computational work. D.W.A., R.D.D., and J.R.W. performed the time-lapse microscopy measurements. M.S.A., J.R.W. and G.S.S. were responsible for the synthetic biology. D.W.A., R.D.D., C.D.C., J.M.M. and M.L.S. analysed the experimental data. J.M.M., N.F.S. and C.D.C. were responsible for simulations. C.D.C. and M.L.S. developed the frequency domain analytical approach. M.L.S. was responsible for integration of experimental, analytical and computational components.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.L.S. (simpsonML1@ornl.gov).

LETTERS

Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus

Wen Jiang^{1,†}, Juan Chang^{1,2}, Joanita Jakana¹, Peter Weigele³, Jonathan King³ & Wah Chiu^{1,2}

The critical viral components for packaging DNA, recognizing and binding to host cells, and injecting the condensed DNA into the host are organized at a single vertex of many icosahedral viruses. These component structures do not share icosahedral symmetry and cannot be resolved using a conventional icosahedral averaging method. Here we report the structure of the entire infectious *Salmonella* bacteriophage epsilon15 (ref. 1) determined from single-particle cryo-electron microscopy, without icosahedral averaging. This structure displays not only the icosahedral shell of 60 hexamers and 11 pentamers, but also the non-icosahedral components at one pentameric vertex. The densities at this vertex can be identified as the 12-subunit portal complex sandwiched between an internal cylindrical core and an external tail hub connecting to six projecting trimeric tailspikes. The viral genome is packed as coaxial coils in at least three outer layers with ~90 terminal nucleotides extending through the protein core and the portal complex and poised for injection. The shell protein from icosahedral reconstruction at higher resolution exhibits a similar fold to that of other double-stranded DNA viruses including herpesvirus^{2–6}, suggesting a common ancestor among these diverse viruses. The image reconstruction approach should be applicable to studying other biological nanomachines with components of mixed symmetries.

Double-stranded DNA (dsDNA) bacteriophages are vectors for gene transfer among enteric bacteria, including important human pathogens⁷. For all the well-studied tailed dsDNA bacteriophages, a preformed procapsid shell is assembled, and the DNA is pumped into the shell through a portal complex located at a single vertex⁸. The bacteriophage tails are also assembled at this vertex. The portal complex together with packaging enzymes have been shown to function as components of a very powerful molecular motor⁹, but it has not been possible to visualize the complex within the intact virion.

Epsilon15 is a short-tailed dsDNA bacteriophage that infects *Salmonella anatum*. Its genome (NCBI accession number NC_004775) contains 39,671 base pairs with 49 open reading frames (Supplementary Fig. 1). Six gene products, coding for structural proteins, were resolved by SDS–polyacrylamide gel electrophoresis (PAGE) and identified by tryptic-digest/mass spectrometry (Supplementary Fig. 2).

Cryo-electron microscopy of frozen hydrated epsilon15 shows intact particles consisting of isometric heads and a protruding density (Fig. 1a). Approximately 15,000 particle images were used for image reconstruction without symmetry imposition. The 20 Å resolution density map of the whole bacteriophage is shown in Fig. 1b (Supplementary Movie 1). The angular-shaped capsid has a $T = 7$

shell, with a larger diameter (700 Å) along the icosahedral five-fold axes and a smaller diameter (650 Å) along the three-fold axes. The virion capsid is made up of 11 pentons and 60 hexons.

One of the five-fold vertices is occupied by six tailspikes (red) surrounding an external tail hub (yellow). A cross-section of the reconstruction (Fig. 1c) and cut-away view (Fig. 1d) reveal, inside the capsid shell (dark green), the densities of the portal complex (purple), a protein core (light green) and the packaged viral DNA (blue). Figure 1e shows three well-resolved concentric shells representing the dsDNA packed coaxially around the five-fold axis.

The ~40-kilobase (kb) epsilon15 dsDNA genome has a length of 14 µm in an extended form and needs to assume a compact arrangement when it is packaged into the capsid. This compact arrangement must also be topologically organized to facilitate efficient dsDNA release during infection. The epsilon15 genome structure is sufficiently resolved to see the individual dsDNA strands with a ~25 Å separation for at least the three outermost layers (Fig. 1d, e), which are packed coaxially around the vertical axis in Fig. 1c, d, coincident with the capsid five-fold/portal/tail axis. Neighbouring dsDNA strands are packed hexagonally, which appears to be the most space-efficient and energetically favourable packing pattern for the cylindrically shaped dsDNA strands. The coaxial dsDNA rings are better resolved in the outermost layers closest to the capsid shell than those in the inner layers (Fig. 1c, d). This suggests that the dsDNA packaging starts from the inner surface of the capsid shell (light blue in Fig. 1d, e), spools towards the inner radius region and ends as a terminal shaft in the portal channel (dark blue in Fig. 1d, e). This type of packing agrees qualitatively with the coaxial spooling model based on scattering¹⁰, simulations¹¹ and previous cryo-electron microscopy reconstruction¹². Our observed spool axis is consistent with the T7 spooling model¹³, but orthogonal to an earlier model of T4 (ref. 10), and does not appear to vary in different layers¹¹.

On the surface of the capsid, six tailspikes, composed of gp20, connect to a central tail hub, which extends out from one of the capsid's five-fold vertices (Figs 1b and 2a). This symmetry mismatch with a six-fold tail complex located at a five-fold vertex is a characteristic feature of tailed dsDNA bacteriophages. Each tailspike consists of two structural domains separated by a bend: a capsid proximal stem bound to the central tail hub and a distal flower-like domain composed of three petals around a central knob (Fig. 2a; see also Supplementary Fig. 4b). The distal end of the flower-like domain is clearly trimeric, suggesting a homo-trimer of gp20. These tailspikes bind and cleave the O-antigen component of the host's cell surface lipopolysaccharide¹. The petal-like densities are candidates for this enzymatic function. The bend may represent a functional hinge of

¹National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, ²Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA. ³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA.

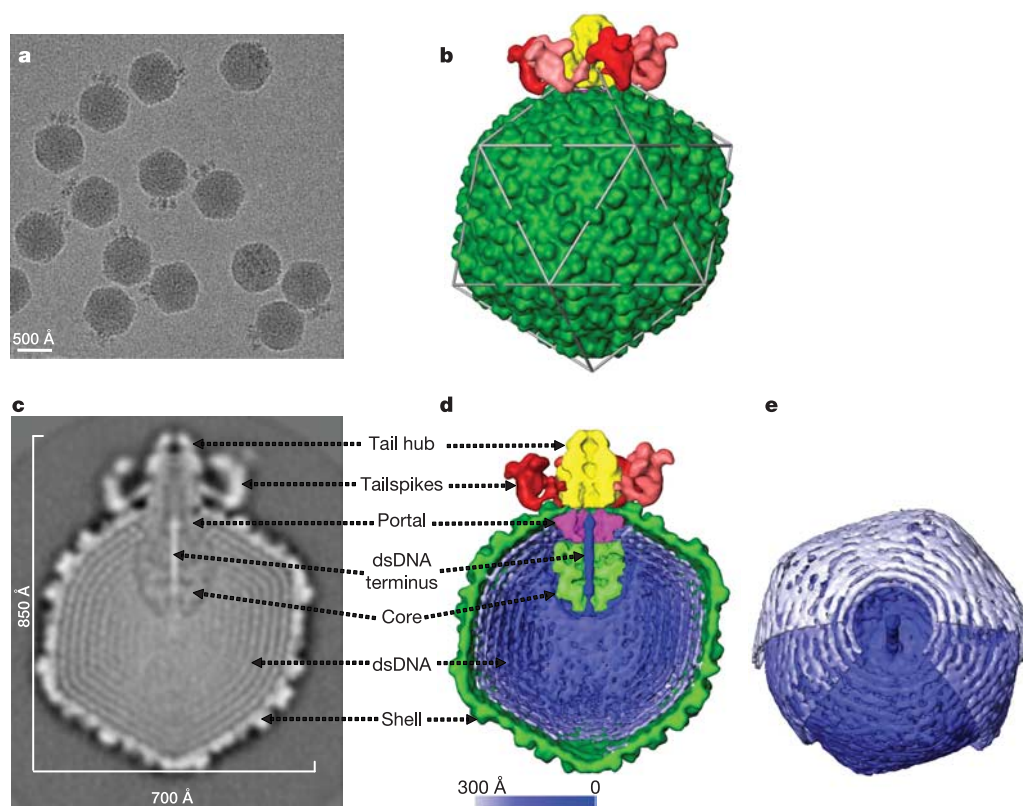


Figure 1 | Structure of epsilon15 bacteriophage. **a**, 200-kV CCD image of epsilon15 particles embedded in vitreous ice. **b**, Surface rendering of the 20 Å resolution three-dimensional map of entire epsilon15 bacteriophage reconstructed without symmetry imposition. The capsid (dark green) exhibits good icosahedral symmetry as indicated by the icosahedral lattice (grey). **c**, **d**, The structural components of epsilon15 bacteriophage are annotated in the central section density (**c**) and the cut-away surface view (**d**) of the three-dimensional density map. Each of the structural components is coloured differently and the same colour scheme is used in Figs 1–3. **e**, A slightly tilted view of the coaxially packed dsDNA genome. Only portions of the first and second layers of dsDNA are shown so that the three outermost layers can be viewed. Colour gradient (light blue to darker blue) is used in **d** and **e** to indicate the likely packaging order of the dsDNA starting from the outer layer towards inner layers and ending as the central straight segment of terminus.

the tailspike that transmits a recognition signal to activate the tail hub for viral genome injection.

The tail deviates from exact six-fold azimuthal symmetry (Fig. 2a, c; see also Supplementary Fig. 4a). Each of the tailspikes has a slightly different orientation (Supplementary Fig. 4a) but similar

conformation, except for tailspike 3 (Supplementary Fig. 4b). These differences among individual tailspikes may arise from the different interaction sites on the capsid shell surface and the symmetry mismatches between the strict five-fold shell and the quasi six-fold tail. Two neighbouring tailspikes (labelled with an asterisk in Fig. 2b)

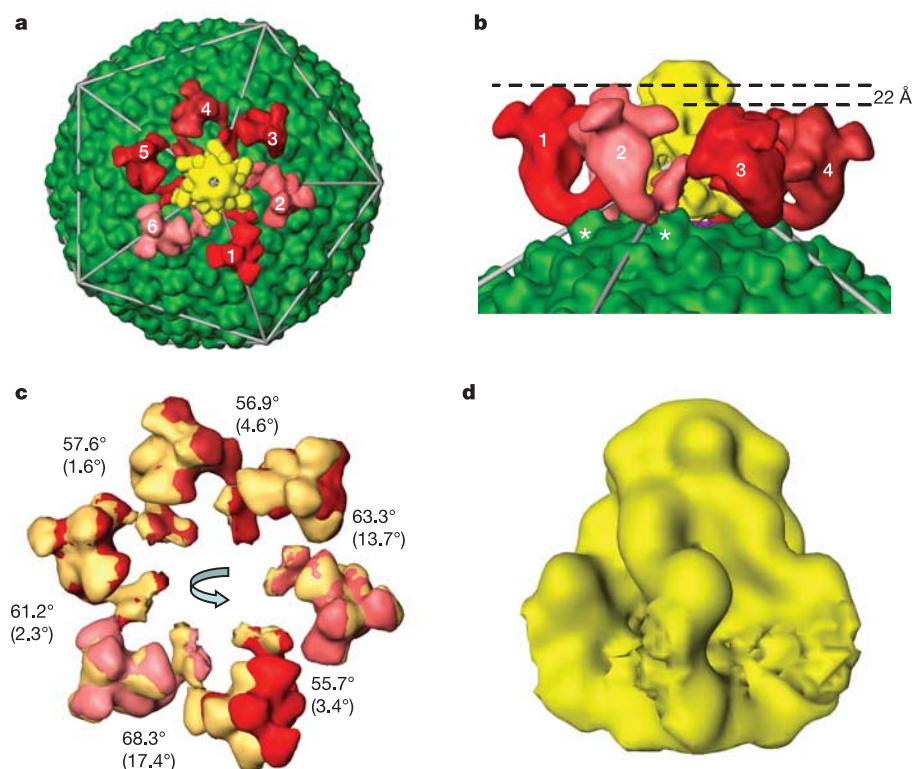


Figure 2 | The tail structure. **a**, Top view of the tail in the epsilon15 reconstruction. Each of the six tailspikes is uniquely coloured with an arbitrary shade of red in order to illustrate the inexact six-fold arrangement around the central tail hub (yellow) with good six-fold symmetry. **b**, Higher-magnification side view of the tailspikes. The contact sites for two (labelled as 1 and 2) out of the six tailspikes are on the capsid surface protrusion densities at local two-fold positions (labelled with an asterisk). **c**, Each of the tailspikes is aligned with its neighbour (anticlockwise) with indicated amount of rotation around an axis that is tilted away from the six-fold axis in different degrees, as indicated in parentheses. The tailspikes in the original positions are displayed in the same shades of red as in **a**, whereas the rotated tailspikes in the new positions are displayed in yellow. **d**, Side view of the tail hub. The segmentation of the tail hub at the interacting regions with the tailspikes is arbitrary because of their tight binding and limited resolution.

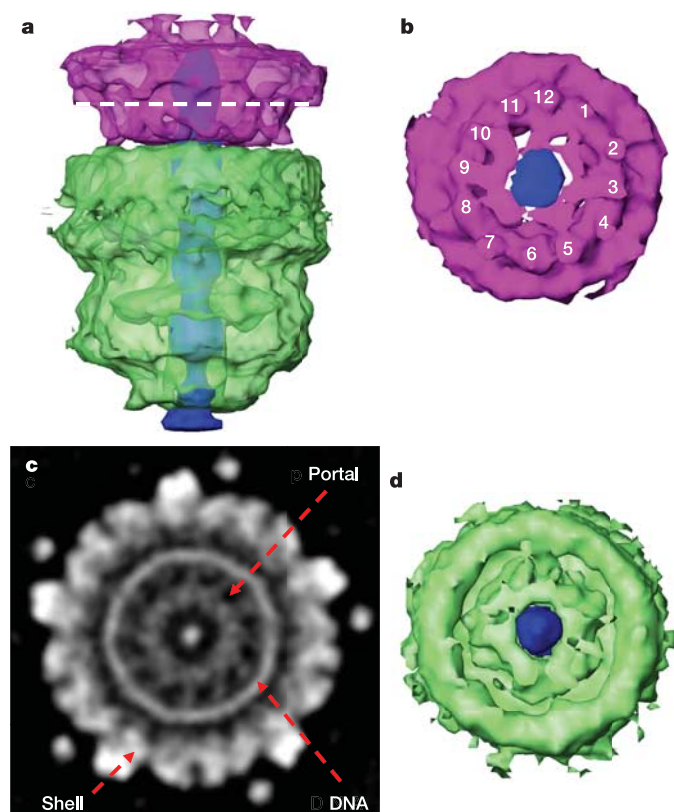


Figure 3 | Structure of the portal complex and internal core. **a**, Side view of the portal complex (pink), internal core (light green) and the putative straight dsDNA terminus (dark blue). The portal complex and internal core are coloured semi-transparently to show that their central channels are filled with the putative dsDNA terminus. **b**, Bottom view of the portal complex with the nearly 12-fold structural features labelled around the filled central channel. **c**, The apparent 12-fold arrangement of densities for the portal complex in a section image of the three-dimensional map. The location of the section is indicated by the dashed line in **a**. **d**, Top view of the internal core (light green) showing the central channel filled with the putative dsDNA terminus (dark blue).

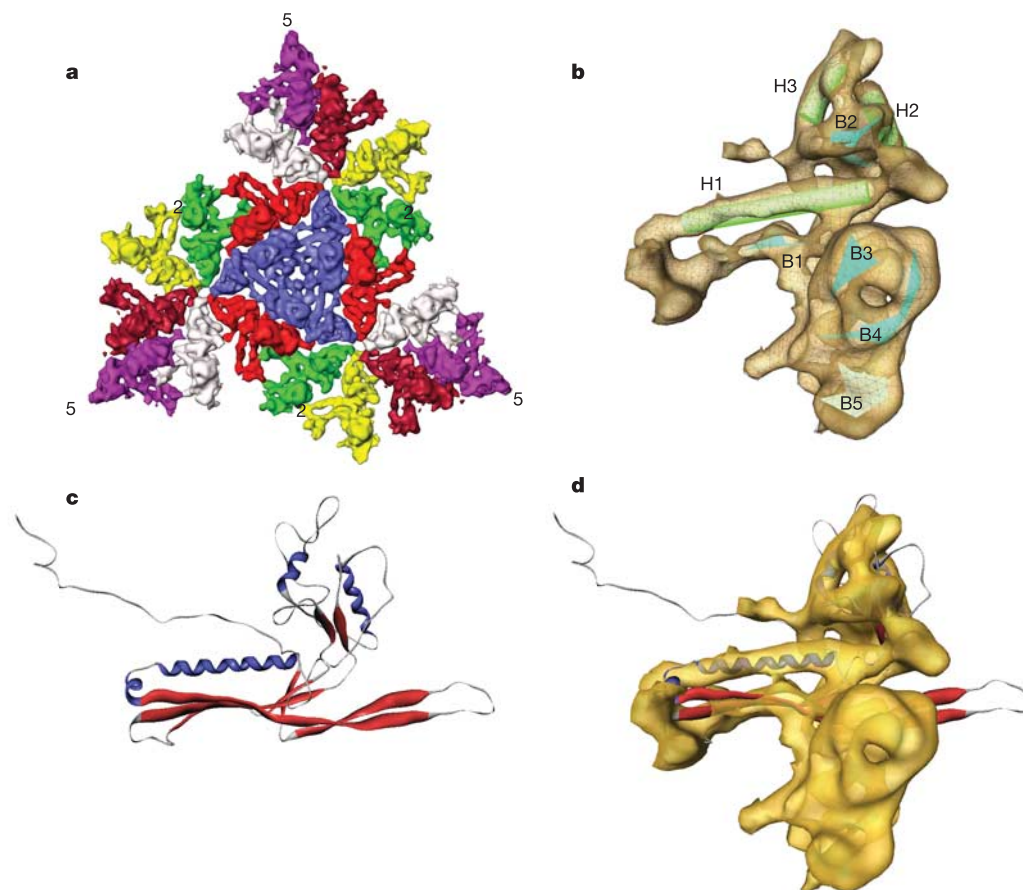


Figure 4 | Shell protein structure. **a**, Surface rendering of three asymmetric units forming one icosahedral face of the $T = 7I$ shell. Each of the seven subunits is coloured differently. **b**, A single averaged shell protein subunit with three helices (H1–H3) and five sheets (B1–B5) annotated. **c**, Ribbon

representation of the atomic model of HK97 head protein gp5 (Protein Data Bank accession 1OHG). Helices are coloured in blue whereas sheets are in red. **d**, Superposition of epsilon15 averaged shell protein subunit and the HK97 head protein gp5.

contact the protrusions at the local two-fold positions of the shell, whereas the other four spikes have tenuous interactions with the shell at different locations at smaller radii. The two elevated contact points force these two tailspikes to a higher radius by 22 Å than the other four tailspikes, and result in a slight overall tilt of the tail (Fig. 2b).

At the centre of the six tailspikes is the tail hub of size ~170 Å (height) by 140 Å (width) (Fig. 2d). The tail hub has good six-fold symmetry (Fig. 2a). It appears to anchor the tailspikes as well as cap the central opening of the portal to prevent premature DNA leakage. Upon binding to the surface of the host cell by the distal petal region of the tailspikes, the induced conformational changes might trigger the opening of the capped central channel (Fig. 1c, d), allowing for DNA injection into the host cell in a manner analogous to a gated ion channel.

Just below the tail hub lays the portal. Although the oligomeric state of portals within intact bacteriophages has long been thought to be 12, it has not been measured until now. Our epsilon15 density map provides unambiguous structural evidence that there are 12 densities at the portal *in situ* (Figs 1c, d and 3a–c; see also Supplementary Fig. 5a, b), which is consistent with previous studies of portal complexes biochemically extracted from bacteriophage particles^{14,15}.

Circumscribed by a well-resolved dsDNA ring, the portal has the appearance of 12 well-resolved turbine-like densities (180 Å in diameter) with nearly equivalent angular spacing (Fig. 3c; see also Supplementary Fig. 5a). This overall feature resembles that seen in other isolated portals with negligible sequence similarity^{16–19}. The azimuthal density distribution of the epsilon15 portal deviates noticeably from perfect 12-fold symmetry (Fig. 3c; see also Supplementary Fig. 5a). The asymmetry of the portal ring may be caused by the non-equivalent interactions between the twelve portal subunits and the five surrounding shell protein subunits.

These observations require that the portal complex has the same spatial relationship with the tailspikes, tail hub and the shell, from particle to particle, implying that there are specific interactions among these components. Otherwise, the portal structure would not be resolved by averaging ~15,000 varying bacteriophage particles.

Figure 3a, d shows an internal density (light green) organized immediately internal to the portal complex. This feature is ~200 Å in height and ~180 Å in width without obvious symmetry. We interpret this to be a protein core reminiscent of the bacteriophage T7 internal eight-fold-symmetric structure observed in electron micrographs²⁰. The lack of obvious symmetry for the epsilon15 core could be authentic or result from an azimuthal variation of the core among particles. The epsilon15 capsid volume can accommodate up to 90 kb dsDNA. Because the epsilon15 genome is only ~40 kb, there is ample space for a protein core of this size in the capsid chamber. The protein core may facilitate the topological ordering of the dsDNA genome during packaging and/or release as suggested for the T7 core²¹.

A long, straight segment of uniform density (~310 Å in length and ~30 Å in width) extends from near the capsid centre along the five-fold capsid–tail axis. Because these densities are significantly higher than those of the surrounding protein core and portal complex (Fig. 1c, d), we interpret this segment of density to be the ‘last-in’ segment of packaged dsDNA (~90 base pairs). It appears to be poised for release and injection into host cell during infection.

This dsDNA terminus passes through the central opening of the portal and ends slightly beyond the portal to where it is capped by densities closing the tail hub’s central channel (Fig. 1c, d). The observation that epsilon15’s portal channel is filled with a straight segment of putative dsDNA terminus is consistent with the observation that bacteriophage SPPI’s terminal segment of packaged dsDNA is tightly associated with its portal protein²².

Imposing the icosahedral symmetry, a 9.5 Å density map was generated from a separate data set imaged to give a higher resolution reconstruction (Supplementary Movie 2 and Supplementary Figs 3b and 6a, b). This map is sufficiently resolved to delineate the

molecular boundaries of each capsid protein subunit and the secondary structure elements (Fig. 4a, b). In the average subunit map, three helices could be identified (Fig. 4b). Helix H1 is ~40 Å long and extends towards the three-fold axes where three neighbouring capsomeres interact intimately (Fig. 4a; see also Supplementary Fig. 8b). Parallel to helix H1 is a large β -sheet that contributes significantly to the local three-fold interactions. The two shorter helices (H2 and H3) are located at the interfaces of adjacent subunits within the same hexon and are nearly parallel to each other.

The dispositions of the helices and sheets are similar to those observed in other bacteriophages and herpesvirus^{2–6} (Fig. 4c, d; see also Supplementary Fig. 7) despite little sequence identity. The similarity extends beyond the protein fold of individual subunits to the molecular interaction patterns between the subunits of a hexon or penton, and at the local three-fold symmetry axes where three neighbouring capsomeres meet (Supplementary Fig. 8a, b). The similarities of the protein fold and assembly of the capsid shell of these viruses suggest that these viruses share an ancient common ancestor and that the capsid protein fold and molecular interactions used to hold the icosahedral shell intact have been strongly conserved during evolution.

METHODS

See Supplementary Methods for further methodological details.

The epsilon15 bacteriophage particles were purified from infected *Salmonella anatum* bacterial culture using gradient centrifugation. The purified bacteriophage particles were prepared for cryo-electron microscopy by rapid freeze plunging²³. The images (Fig. 1a) for the reconstruction without symmetry imposition were recorded on a Gatan 4kx4k CCD camera using the JAMES imaging system²⁴ attached to a JEM2010F microscope operated at 200 kV with a Gatan liquid nitrogen specimen cryo-holder. The images (Supplementary Fig. 6a) for the icosahedral reconstruction at higher resolution were taken in a JEM3000SFF microscope operated at 300 kV and at liquid helium specimen temperature. These images were recorded on KODAK SO163 films and were digitized using a Nikon Super CoolScan 9000 ED scanner.

In both sets of data, the individual bacteriophage particle images were pre-processed with the standard procedures. The SAVR software package²⁵ was used to reconstruct the three-dimensional maps assuming icosahedral symmetry for both data sets. The icosahedral symmetry was further relaxed to C1 symmetry to generate the reconstruction without symmetry imposition using a set of newly developed programs within the EMAN package²⁶. The 20 Å resolution map without any symmetry imposition was generated using ~15,000 particle images from ~950 CCD frames of the 200-kV data set. The 9.5 Å map with icosahedral symmetry imposition was generated using ~6,000 particle images from ~200 micrographs of the 300-kV data set.

Amira (<http://www.amiravis.com>) visualization software was used for most of the segmentation and graphics displays. Alignment and similarity comparison of the individual subunits of the shell protein were performed using the foldhunter program²⁷. Identification of α -helices and β -sheets was performed using the AIRS program, which provides a graphic interface to the ssehunter program²⁷ in the Chimera visualization software package²⁸.

Received 17 August; accepted 23 November 2005.

- McConnell, M., Reznick, A. & Wright, A. Studies on the initial interactions of bacteriophage epsilon15 with its host cell, *Salmonella anatum*. *Virology* **94**, 10–23 (1979).
- Jiang, W. *et al.* Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nature Struct. Biol.* **10**, 131–135 (2003).
- Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
- Fokine, A. *et al.* Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc. Natl Acad. Sci. USA* **102**, 7163–7168 (2005).
- Morais, M. C. *et al.* Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol. Cell* **18**, 149–159 (2005).
- Baker, M. L., Jiang, W., Rixon, F. J. & Chiu, W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* **79**, 14967–14970 (2005).
- Breitbart, M., Rohwer, F. & Abedon, S. T. in *Phages: their Role in Bacterial Pathogenesis and Biotechnology* (eds Waldor, M. K., Friedman, D. I. & Adhya, S. L.) 66–91 (ASM Press, Washington DC, 2005).

8. Bazinet, C. & King, J. The DNA translocating vertex of dsDNA bacteriophage. *Annu. Rev. Microbiol.* **39**, 109–129 (1985).
9. Smith, D. E. *et al.* The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **413**, 748–752 (2001).
10. Earnshaw, W. C., King, J., Harrison, S. C. & Eiserling, F. A. The structural organization of DNA packaged within the heads of T4 wild-type, isometric and giant bacteriophages. *Cell* **14**, 559–568 (1978).
11. Arsuaga, J., Tan, R. K., Vazquez, M., Summers de, W. & Harvey, S. C. Investigation of viral DNA packaging using molecular mechanics models. *Biophys. Chem.* **101–102**, 475–484 (2002).
12. Zhang, Z. *et al.* Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy. *J. Mol. Biol.* **297**, 615–626 (2000).
13. Cerritelli, M. E. *et al.* Encapsidated conformation of bacteriophage T7 DNA. *Cell* **91**, 271–280 (1997).
14. Lurz, R. *et al.* Structural organisation of the head-to-tail interface of a bacterial virus. *J. Mol. Biol.* **310**, 1027–1037 (2001).
15. Carazo, J. M., Fujisawa, H., Nakasu, S. & Carrascosa, J. L. Bacteriophage T3 gene 8 product oligomer structure. *J. Ultrastruct. Mol. Struct. Res.* **94**, 105–113 (1986).
16. Simpson, A. A. *et al.* Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408**, 745–750 (2000).
17. Orlova, E. V. *et al.* Structure of a viral DNA gatekeeper at 10 Å resolution by cryo-electron microscopy. *EMBO J.* **22**, 1255–1262 (2003).
18. Agirrezabala, X. *et al.* Structure of the connector of bacteriophage T7 at 8 Å resolution: structural homologies of a basic component of a DNA translocating machinery. *J. Mol. Biol.* **347**, 895–902 (2005).
19. Tang, L., Marion, W. R., Cingolani, G., Prevelige, P. E. & Johnson, J. E. Three-dimensional structure of the bacteriophage P22 tail machine. *EMBO J.* **24**, 2087–2095 (2005).
20. Cerritelli, M. E. *et al.* A second symmetry mismatch at the portal vertex of bacteriophage T7: 8-fold symmetry in the procapsid core. *J. Mol. Biol.* **327**, 1–6 (2003).
21. Molineux, I. J. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol. Microbiol.* **40**, 1–8 (2001).
22. Tavares, P., Lurz, R., Stiege, A., Ruckert, B. & Trautner, T. A. Sequential headful packaging and fate of the cleaved DNA ends in bacteriophage SPP1. *J. Mol. Biol.* **264**, 954–967 (1996).
23. Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).
24. Booth, C. R. *et al.* A 9 Å single particle reconstruction from CCD captured images on a 200 kV electron cryomicroscope. *J. Struct. Biol.* **147**, 116–127 (2004).
25. Jiang, W. *et al.* Semi-automated icosahedral particle reconstruction at sub-nanometer resolution. *J. Struct. Biol.* **136**, 214–225 (2001).
26. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
27. Jiang, W., Baker, M. L., Ludtke, S. J. & Chiu, W. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.* **308**, 1033–1044 (2001).
28. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the support of grants from National Institutes of Health and the Robert Welch Foundation. We thank M. Dougherty for the production of the animations, M. Baker for the AIRS program for secondary structure element identification, and M. F. Schmid and F. Rixon for discussions.

Author Contributions W.J. developed the image processing methods and solved and analysed the structures; J.C. and J.J. collected the 200- and 300-kV image data respectively; P.W. did the biochemical preparation and analysis; and W.J., P.W., J.K. and W.C. interpreted the structure and wrote the manuscript.

Author Information The three-dimensional density maps have been deposited into the EBI-MSD EMD database with accession codes EMD-1175 for the complete structure without symmetry imposition and EMD-1176 for the icosahedral shell structure. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.C. (wah@bcm.edu).

The mechanism of DNA replication primer synthesis by RNA polymerase

Nikolay Zenkin^{1,3}, Tatyana Naryshkina¹, Konstantin Kuznedelov¹ & Konstantin Severinov^{1,2,3}

RNA primers for DNA replication are usually synthesized by specialized enzymes, the primases¹. However, some replication systems have evolved to use cellular DNA-dependent RNA polymerase for primer synthesis^{1,2}. The main requirement for the replication primer, an exposed RNA 3' end annealed to the DNA template, is not compatible with known conformations of the transcription elongation complex³, raising a question of how the priming is achieved. Here we show that a previously unrecognized kind of transcription complex is formed during RNA polymerase-catalysed synthesis of the M13 bacteriophage replication primer. The complex contains an overextended RNA–DNA hybrid bound in the RNA–polymerase trough that is normally occupied by downstream double-stranded DNA, thus leaving the 3' end of the RNA available for interaction with DNA polymerase. Transcription complexes with similar topology may prime the replication of other bacterial mobile elements and may regulate transcription elongation under conditions that favour the formation of an extended RNA–DNA hybrid.

A fundamental difference between RNA polymerase (RNAP) and DNA polymerase (DNAP) is that the latter requires a primer to initiate template-dependent nucleic acid synthesis, whereas the former does not. Several viral (T4 and M13) and plasmid (ColEI and F) replicons use bacterial RNAP to synthesize a replication primer. The mechanism(s) that ensure the transfer of the primer's 3' end, which must be annealed to the template DNA, from RNAP to DNAP are not known.

Here we characterize primer synthesis from the minus-strand origin (*ori*) of bacteriophage M13. *Escherichia coli* RNAP σ^{70} holoenzyme ($E\sigma^{70}$) specifically recognizes an imperfect hairpin formed by the *ori* of bacteriophage M13 (Fig. 1a) and synthesizes a transcript 18–20 nucleotides (nt) long (primer RNA; pRNA) that serves as a primer for DNAP III (refs 2, 4, 5). The cessation of RNA synthesis during the normal transcription cycle can occur either because of backtracking, when RNAP moves backward, extruding the 3' end of the nascent RNA into the enzyme's secondary channel while maintaining the normal length of the RNA–DNA hybrid (eight or nine base pairs), or because of transcription termination, when the transcription elongation complex (EC) falls apart. An *in vitro* pRNA synthesis reaction was performed on *ori* DNA⁴; the product was fractionated on a gel-filtration column, and fractions were analysed for the presence of radioactively labelled pRNA and $E\sigma^{70}$. $E\sigma^{70}$ and pRNA were eluted together from the column (Fig. 1b), whereas pRNA from deproteinized reaction was eluted much later (indicated by a downward arrow in Fig. 1b). pRNA in column fractions was sensitive to RNase H (Fig. 1b), indicating that *ori* DNA was also eluted with $E\sigma^{70}$ and pRNA. Thus, pRNA is not produced by a normal transcription termination event, which should have resulted in release of the nascent RNA from the transcription complex. Instead, a ternary priming complex is formed. The formation of

the priming complex was confirmed in experiments with immobilized RNAP (Fig. 1c) and by crosslinking RNAP β and σ^{70} subunits and *ori* DNA to derivatized 5' end of the nascent pRNA (Fig. 1d).

The priming complex was compared with two well-characterized elongation complexes formed on a T7 A1 promoter-containing DNA: an active elongation complex containing a 20-nt RNA (EC20), and backtracked EC27 complex⁶. The complexes were prepared with the use of immobilized RNAP and were washed away from unincorporated substrates⁶. Unlike EC20, but similarly to EC27, the priming complex was unable to extend the nascent RNA even in the presence of high concentrations of NTPs (Fig. 1e; compare lanes 2, 7 and 12) or undergo processive pyrophosphorylation (Fig. 1e; compare lanes 4, 9 and 14). However, unlike the backtracked complex (and similarly to the active complex), the priming complex was resistant to GreB (Fig. 1e; compare lanes 3, 8 and 13). The priming complex possessed a unique property: it was sensitive to the addition of RNase H, which digested pRNA (Fig. 1e, lane 15). In contrast, the nascent RNAs in EC20 and EC27 were fully resistant to digestion with RNase H (Fig. 1e, lanes 5 and 10, respectively). pRNA in the priming complex was also sensitive to *E. coli* exonuclease III, an enzyme that specifically degrades double-stranded nucleic acids from exposed 3' ends (data not shown). The susceptibility of pRNA to RNase H and exonuclease III (ExoIII) indicates that the 3' end of the pRNA transcript is annealed to the *ori* template and is exposed. Thus, the ternary priming complex is distinct from known transcription complexes.

In contrast to the situation during normal transcript elongation, no upstream duplex DNA should form during pRNA synthesis (Fig. 1a). Elongation of nascent pRNA might therefore be prevented as a result of the formation of overextended (compared with the eight or nine base pairs found in normal EC) hybrid between the nascent pRNA and the *ori* DNA. This hypothesis predicts that decreasing the strength of pRNA interaction with transcribed portion of *ori* DNA should decrease the amount of pRNA synthesized, whereas the amount of the read-through transcript should increase. To weaken the interaction, we performed pRNA synthesis in the presence of ITP instead of GTP (Fig. 2a). ITP suppressed pRNA synthesis; instead, a longer RNA that resulted from transcription to the end of the 50-nt *ori* template was present (Fig. 2a, lane 3). Thus, the strength of base-pairing between RNA and *ori* DNA controls pRNA production.

Three consecutive GMP residues close to the 5' end of pRNA might be particularly important for blocking the disengagement of the 5' end of the pRNA from the template DNA. When a mutant *ori* template in which the three G–C base pairs at positions +2, +3 and +4 had been replaced by T–A pairs was used for pRNA synthesis, no pRNA was synthesized and only the full-size transcript was present (Fig. 2a; compare lanes 1 and 2). However, when 5-BrUTP, which forms three Watson–Crick bonds with adenosine, was used instead of UTP, pRNA synthesis on the mutant template was partly restored

¹Waksman Institute, ²Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA. ³Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182 Russia.

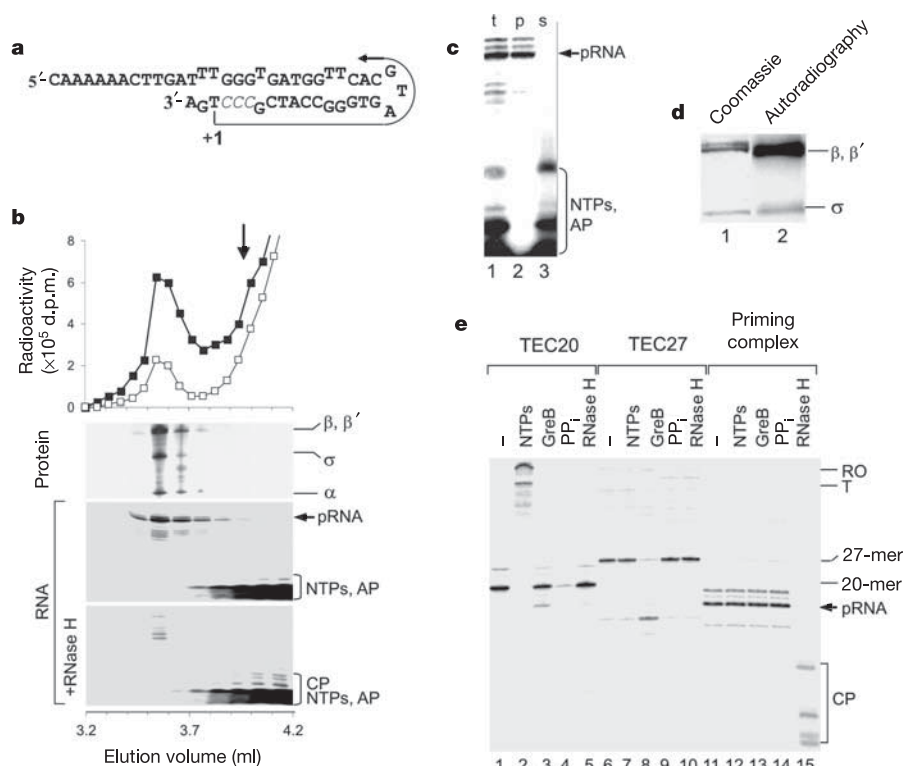


Figure 1 | Properties of the priming complex. **a**, Structure of the M13 origin of replication. pRNA is shown as an arrow. dCMPs that are changed to dAMPs in the mutant template (Fig. 2a) are in italic. **b**, A gel-filtration elution profile of radioactive pRNA synthesis reactions not treated (black squares) and treated (empty squares) with RNase H. Fractions were analysed by SDS-PAGE for protein content and by urea-PAGE for pRNA. In the lower panel, fractions were treated with RNase H. A black arrow in the plot indicates the elution volume of pRNA from deproteinized pRNA synthesis reaction. AP, abortive products; CP, cleavage products. **c**, pRNA was

synthesized using RNAP immobilized on Ni-NTA agarose. The bead suspension (t) was separated into a pellet (p) and the supernatant (s) and the contents were resolved by urea-PAGE. **d**, pRNA containing a crosslinkable group at its 5' end was synthesized, the crosslink was induced and reaction products were separated by SDS-PAGE and revealed by protein staining and autoradiography. **e**, Comparison of the properties of the priming complex (lanes 11–15), the active elongation complex (TEC20, lanes 1–5), and the backtracked elongation complex (TEC27, lanes 6–10). RO, run-off; T, terminator. PP_i, pyrophosphate.

(data not shown). Thus, the strength of base-pairing between the 5' end-proximal RNA residues and the template DNA controls the formation of the defined 3' end of pRNA.

To show directly that the 5' end of pRNA stays annealed to the *ori* template, chemical crosslinking with artificially stalled *ori* complexes containing 8, 11, 13 or 15-nt nascent RNAs was performed (Fig. 2b). The 5' ends of nascent RNAs contained an activatable alkylating group that can crosslink to DNA (provided that an RNA–DNA hybrid exists). Crosslinking of complexes with 8-nt IMP-containing RNA (Fig. 2b, lane 10), resulted in a radioactive band that was sensitive to treatment with DNase I (data not shown) and therefore corresponded to RNA–DNA crosslinking. Little or no RNA–DNA crosslinking was observed in complexes with IMP-containing RNA longer than 8 nt (Fig. 2b, lanes 12, 14 and 16), indicating that the RNA–DNA hybrid is eight or nine base pairs long and that the 5' end of longer RNAs is directed into the RNA exit channel, as expected for 'normal' elongation. Crosslinking of complexes with GMP-containing RNA revealed the presence of an RNA–DNA crosslink for all complexes tested (Fig. 2b, lanes 2, 4, 6 and 8). The result strongly indicates that, during priming complex formation, the 5' end of the RNA fails to disengage from the template strand, which results in the formation of an overextended RNA–DNA hybrid.

The RNA–DNA hybrid length in the normal transcription elongation complex is thought to be determined by several structural elements of the RNAP β' subunit—the rudder, the lid and the zipper. These elements abut the upstream edge of the hybrid and may direct the nascent RNA to the exit channel. To determine which of these elements contributes to the formation of pRNA, mutant RNAPs lacking the rudder, the lid or the zipper were used for pRNA

synthesis. No defects in pRNA synthesis were observed with enzymes lacking the rudder or the zipper (Fig. 2c, lanes 3 and 4). However, the enzyme lacking the lid did not produce pRNA: it transcribed to the end of the template (Fig. 2c, lane 2).

The data presented so far are consistent with an idea that in the absence of a non-template strand, a collision of the 5' end of the nascent RNA with the β' lid leads to the formation of overextended hybrid. The RNA synthesis continues for some time but is ultimately stopped when a critical hybrid length is reached. One can therefore expect that the initially active elongation complexes should start resembling the priming complex as the length of the nascent RNA increases beyond eight or nine bases. We compared artificially stalled *ori* DNA complexes containing RNAs 7, 10, 12 and 14 nt long. EC7, EC10 and EC12 responded to the addition of NTPs or pyrophosphate by, respectively, almost quantitative extension (Fig. 2d, lanes 2, 6 and 10) or shortening (Fig. 2d, lanes 3, 7, and 11) of original transcripts, whereas EC14 was resistant to both treatments (Fig. 2d, lanes 14 and 15). EC7 and EC10 were resistant to RNase H (Fig. 2d, lanes 4 and 8). In contrast, EC12 and EC14 were sensitive to RNase H (Fig. 2d, lanes 12 and 16). Thus, EC12 can be regarded as an intermediate (or a mixture) between active elongation complex and the priming complex.

We next defined the trajectory of pRNA in the priming complex. The position of the 5' end of the nascent pRNA was determined by mapping a crosslink with the RNAP β subunit (Fig. 3a). Mapping positioned the 5' end of pRNA close to, or in contact with, β Lys 1065, a conserved residue located close to the 3' end of the nascent RNA in normal elongation complexes^{2,7}. Thus, after pRNA synthesis has been completed, the 5' end of the nascent transcript, which should have

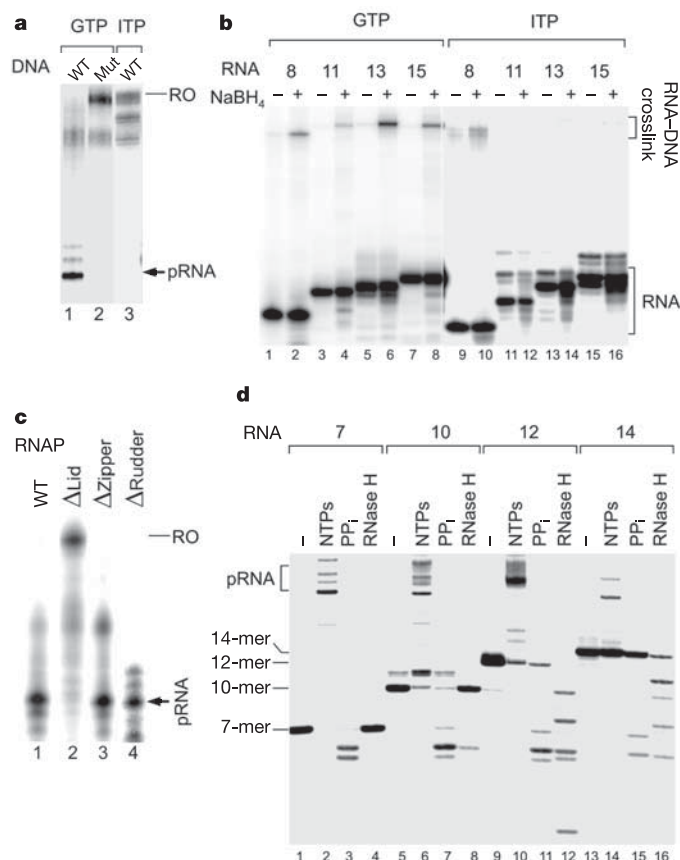


Figure 2 | Overextended RNA-DNA hybrid causes formation of the priming complex. **a**, pRNA synthesis was performed on the wild-type (wt, lanes 1 and 3) origin fragment or a mutant fragment containing dAMPs at positions +2, +3 and +4 of the template strand instead of dCMP (Mut, lane 2) in the presence of GTP (lanes 1 and 2) or ITP (lane 3). RO, run-off. **b**, The RNA-DNA crosslink from the 5' end of the RNA in stalled complexes, obtained as described in Methods, with RNAs of the indicated lengths, containing either GMP (lanes 1–8) or IMP (lanes 9–16). **c**, pRNA synthesis by wild-type RNAP (lane 1) and by mutant RNAPs lacking the lid (Δ Lid, lane 2), the zipper (Δ Zipper, lane 3) or the rudder (Δ Rudder, lane 4). **d**, Stepwise formation of the priming complex. Stalled complexes containing the indicated nascent RNAs were obtained as described in Methods. Complexes were analysed for their ability to extend RNA after the addition of NTPs (lanes 2, 6, 10 and 14) and to undergo pyrophosphorolysis (lanes 3, 7, 11 and 15) and for the susceptibility of the RNA to treatment with RNase H (lanes 4, 8, 12 and 16).

been 18 nt upstream of Lys 1065 during the synthesis of the last phosphodiester bond, moves back to its vicinity.

The ExoIII and crosslinking analysis argues that, in the priming complex, pRNA is annealed to the *ori* DNA throughout its length. To localize the position of the RNA-DNA hybrid within the priming complex, we used bacteriophage T7 gp2, a small protein that binds to the β' downstream jaw domain⁸. The β' jaw is part of a trough that interacts with double-stranded DNA downstream of the catalytic centre; T7 gp2 prevents this interaction. The addition of T7 gp2 to the priming complex resulted in the release of pRNA, which was sensitive to RNase H, indicating that it remained annealed to the minus-strand origin template (Fig. 3b). Thus, the β' jaw is required for the priming complex stability, and the extended RNA-DNA hybrid must be bound in the RNAP downstream trough, with the 5' end of pRNA near the RNAP active centre and the 3' end annealed to the DNA template and available for DNAP (see Supplementary Fig. 1). It is tempting to speculate that direct contacts between RNAP in the priming complex and components of DNA replication machinery might further improve the efficiency of primer recognition.

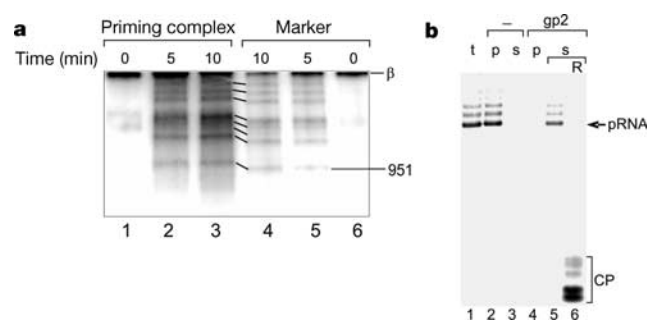


Figure 3 | Structure of the priming complex. **a**, Mapping of the crosslink from the 5' end of pRNA. The β subunit radioactively labelled by crosslinking to the 5' end of pRNA in the priming complex (lanes 1–3) was treated with CNBr under single-hit conditions for different durations and the cleavage products were separated by SDS-PAGE. As a marker, the β subunit affinity labelled on Lys 1065 (ref. 15) and subjected to identical CNBr treatment was used (lanes 4–6). Solid lines connect corresponding cleavage products from the 10-min lanes. **b**, T7 gp2 destroys the priming complex. pRNA synthesis reactions were performed on the solid phase. After incubation with or without gp2, the reaction mixtures (t, lane 1) were divided into supernatant (s, lanes 3, 5 and 6) and pellet (p, lanes 2 and 4); in lane 6 the supernatant was treated with RNase H. CP, cleavage products.

Our principal finding is the demonstration that the formation of an overextended RNA-DNA hybrid leads to the cessation of transcription and the rearrangement of the RNAP elongation complex into a complex with highly unusual topology that satisfies the requirements for priming of DNA synthesis. The sequence of events during pRNA formation is shown schematically in Supplementary Fig. 1. In brief, the absence of upstream DNA duplex formation allows overextended RNA-DNA hybrid to form. A clash with the β' lid upon RNA growth beyond 8 or 9 nt alters the trajectory of the hybrid and leads to increased strain. Once the RNA-DNA hybrid reaches a certain critical length, the RNAP catalytic centre disengages from the 3' end of the RNA and the enzyme slides backwards along the RNA-DNA hybrid forming the priming complex. The backsliding differs from the previously described RNAP backtracking that results in the 3'-end-proximal portion of the RNA being threaded through the RNAP secondary channel^{3,6,7}.

Filamentous phages⁵ as well as many drug resistance plasmids from Gram-positive and Gram-negative bacteria^{9–11} have *ori* structures that are similar to the M13 *ori*, indicating that they might use this mechanism of priming. We speculate that a similar mechanism might be responsible for the formation of extensive RNA-DNA hybrids during RNAP-catalysed priming of replication on double-stranded replicons such as ColE1 (ref. 12).

METHODS

Wild-type and mutant RNAPs. *E. coli* holo-RNAP, *Thermus aquaticus* core RNAPs and *T. aquaticus* σ^A were purified as described previously^{13,14}. Mutant *T. aquaticus* RNAPs lacking the zipper or the lid domains were obtained by genetically substituting the β' subunit residues 27–42 (zipper) for a Gly-Gly linker or residues 526–539 (lid) for a Gly linker, with the use of a previously described recombinant *T. aquaticus* RNAP co-expression system¹⁴. Mutant *T. aquaticus* RNAP lacking the rudder domain was described previously¹⁴.

Transcription in vitro. Stalled elongation complexes TEC20 and TEC27 were obtained as described elsewhere⁶. *In vitro* synthesis of pRNA was performed essentially as described⁴. When needed, complexes were isolated on Ni²⁺-nitrilotriacetate agarose (Qiagen) using His₆-tagged RNAP¹³. To obtain the stalled complexes, synthesis on mutant templates (Fig. 2d) was initiated with 50 μ M CpA and 10 μ M CTP, 10 μ M UTP and 10 μ M GTP or 300 μ M ITP. In experiments with IMP-containing stalled complexes, 3'-dATP was incorporated into the 3' end of RNAs to reduce the readthrough (this explains why complexes in Fig. 2b contained RNAs 1 nt longer than those in Fig. 2d). RNase H (3 units), 1 or 10 units of ExoIII, 100 μ M NTPs, 5 mM pyrophosphate or 500 nM GreB were added to the complexes for 10 min at 37°C. For

the competition assay, 500 nM T7gp2 was added to the immobilized complex, reactions were incubated for 10 min at 37 °C and then divided into supernatant and pellet. Reactions were stopped with formamide-containing buffer, and products were separated on 15–23% denaturing PAGE and detected by phosphorimaging.

Gel filtration. Reaction mixtures (50–100 µl) were loaded on a 10-ml Sephacryl S200 column equilibrated with transcription buffer⁴, and 200-µl fractions were collected at a flow rate of 0.5 ml min⁻¹. Fractions were analysed for protein content by SDS-PAGE and for RNA content as above.

Crosslinking. For RNA–DNA crosslinking, stalled complexes were obtained as above, except that Cpa was derivatized with an [N,N-bis-(2-iodoethyl)]amino-benzaldehyde group on the 5' end. Crosslinking was induced by the addition of NaBH₄ to a final concentration of 1 mg ml⁻¹, and products were analysed as above. For RNA–protein crosslinking, pRNA synthesis was initiated with the *o*-formylphenyl ester of AMP, a lysine-specific reagent, and crosslinking was performed and mapped (Fig. 3a) as described^{4,15}.

Received 26 July; accepted 20 October 2005.

- Kornberg, A. & Baker, T. A. *DNA Replication* 2nd edn (Freeman, New York, 1992).
- Geider, K. & Kornberg, A. Conversion of the M13 viral single strand to the double-stranded replicative forms by purified proteins. *J. Biol. Chem.* **249**, 3999–4005 (1974).
- Nudler, E. Transcription elongation: structural basis and mechanisms. *J. Mol. Biol.* **288**, 1–12 (1999).
- Zenkin, N. & Severinov, K. The role of RNA polymerase sigma subunit in promoter-independent initiation of transcription. *Proc. Natl Acad. Sci. USA* **101**, 4396–4400 (2004).
- Horiuchi, K. Initiation mechanisms in replication of filamentous phage DNA. *Genes Cells* **2**, 425–432 (1997).
- Komissarova, N. & Kashlev, M. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA* **94**, 1755–1760 (1997).
- Korzheva, N. *et al.* A structural model of transcription elongation. *Science* **289**, 619–625 (2000).
- Nechaev, S. & Severinov, K. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. *J. Mol. Biol.* **289**, 815–826 (1999).
- Kramer, M. G., Espinosa, M., Misra, T. K. & Khan, S. A. Characterization of a single-strand origin, ssoU, required for broad host range replication of rolling-circle plasmids. *Mol. Microbiol.* **33**, 466–475 (1999).
- Kramer, M. G., Khan, S. A. & Espinosa, M. Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. *EMBO J.* **16**, 5784–5795 (1997).
- Masai, H. & Arai, K. Frp: a novel single-stranded DNA promoter for transcription and for primer RNA synthesis of DNA replication. *Cell* **89**, 897–907 (1997).
- Masukata, H. & Tomizawa, J. A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* **62**, 331–338 (1990).
- Kashlev, M. *et al.* Histidine-tagged RNA polymerase of *Escherichia coli* and transcription in solid phase. *Methods Enzymol.* **274**, 326–334 (1996).
- Kuznedelov, K., Korzheva, N., Mustaev, A. & Severinov, K. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA–DNA hybrid length. *EMBO J.* **21**, 1369–1378 (2002).
- Mustaev, A. *et al.* Strategies and methods of cross-linking of RNA polymerase active center. *Methods Enzymol.* **371**, 191–206 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is dedicated to Dmitry Saloni who inspired it. We thank A. Mustaev for crosslinking reagents. This work was supported by an NIH R01 grant and a Borroughs Wellcome Career Award to K.S.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.Z. (nicserzen@mail.ru) or K.S. (severik@waksman.rutgers.edu).

DNA primase acts as a molecular brake in DNA replication

Jong-Bong Lee¹, Richard K. Hite², Samir M. Hamdan¹, X. Sunney Xie³, Charles C. Richardson¹
& Antoine M. van Oijen¹

A hallmark feature of DNA replication is the coordination between the continuous polymerization of nucleotides on the leading strand and the discontinuous synthesis of DNA on the lagging strand¹. This synchronization requires a precisely timed series of enzymatic steps that control the synthesis of an RNA primer, the recycling of the lagging-strand DNA polymerase, and the production of an Okazaki fragment. Primases synthesize RNA primers at a rate that is orders of magnitude lower^{2–4} than the rate of DNA synthesis by the DNA polymerases at the fork. Furthermore, the recycling of the lagging-strand DNA polymerase from a finished Okazaki fragment to a new primer is inherently slower than the rate of nucleotide polymerization⁵. Different models have been put forward to explain how these slow enzymatic steps can take place at the lagging strand without losing coordination with the continuous and fast leading-strand synthesis^{6–8}. Nonetheless, a clear picture remains elusive. Here we use single-molecule techniques to study the kinetics of a multiprotein replication complex from bacteriophage T7 and to characterize the effect of primase activity on fork progression. We observe the synthesis of primers on the lagging strand to cause transient pausing of the highly processive leading-strand synthesis. In the presence of both leading- and lagging-strand synthesis, we observe the formation and release of a replication loop on the lagging strand. Before loop formation, the primase acts as a molecular brake and transiently halts progression of the replication fork. This observation suggests a mechanism that prevents leading-strand synthesis from outpacing lagging-strand synthesis during the slow enzymatic steps on the lagging strand.

An attractive model system for studying the orchestration at the replication fork is the replication machinery of bacteriophage T7. It can be reconstituted *in vitro* using a small number of purified proteins (Fig. 1a) and its organization closely mimics that of *Escherichia coli* and more complex organisms⁹. The T7 DNA polymerase consists of a 1:1 complex of the T7 gene 5 protein (gp5), encoded by the phage, and the thioredoxin (trx) processivity factor, encoded by the *E. coli* host¹⁰. The T7 gene 4 protein (gp4) assembles as a hexamer and provides both helicase and primase activities^{11,12}. The helicase activity, required for unwinding the parental DNA strands, is located in the carboxy-terminal half, and the primase activity, capable of synthesizing the tetranucleotide primers that are required to initiate lagging-strand DNA synthesis, is located in the amino-terminal half (see Fig. 1a, b). The position of the primase active site on the outside of the hexamer¹³ would, in principle, allow the helicase to continue unwinding while a primer is synthesized on the lagging strand. A replication loop formed on the lagging strand of the replication fork reorients the lagging-strand DNA polymerase so

that it advances in parallel with the leading-strand polymerase.

Here we characterize the kinetics of leading- and lagging-strand synthesis at the single-molecule level by monitoring the length of individual DNA molecules during the replication reaction. The 5' end of one strand of a 48.5 kilobase (kb)-long duplex λ phage DNA molecule is attached to the bottom surface of a glass flow cell using a biotin–streptavidin linker. The opposite 3' end bears a digoxigenin moiety and is linked to a bead that is 2.8 μm in diameter and coated with anti-digoxigenin antibody (Fig. 2a, b). A constant laminar flow is applied above the surface such that the resultant drag on the beads stretches the DNA molecules with a force of 3 pN (ref. 14). This force is well within the regime in which the force reduces the configurational freedom of the DNA but does not influence protein–DNA interactions¹⁵. We measure changes in the lengths of the individual DNA molecules by imaging the beads and tracking their positions.

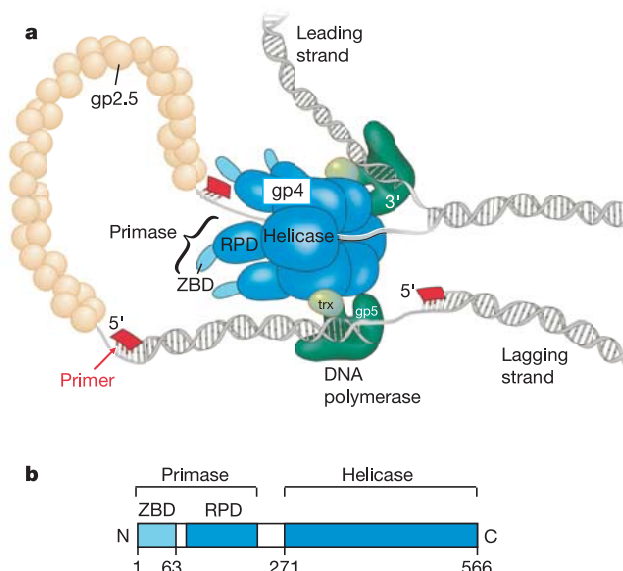


Figure 1 | The T7 replication fork. **a**, The bacteriophage T7 replisome consists of the hexameric T7 gene 4 protein (gp4) and two copies of the T7 DNA polymerase (T7 gene 5 protein (gp5) complexed with *E. coli* thioredoxin (trx)). T7 gene 2.5 protein (gp2.5), the ssDNA-binding protein, coats the transiently exposed ssDNA in the replication loop. **b**, Gp4 consists of a primase and a helicase domain, connected by a linker region. The primase domain consists of two subdomains: a zinc-binding domain (ZBD) and the RNA polymerase domain (RPD).

¹Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, 250 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Harvard Medical School, Graduate Program of Biological and Biomedical Sciences, 220 Longwood Avenue, Boston, Massachusetts 02115, USA. ³Harvard University, Department of Chemistry and Chemical Biology, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.

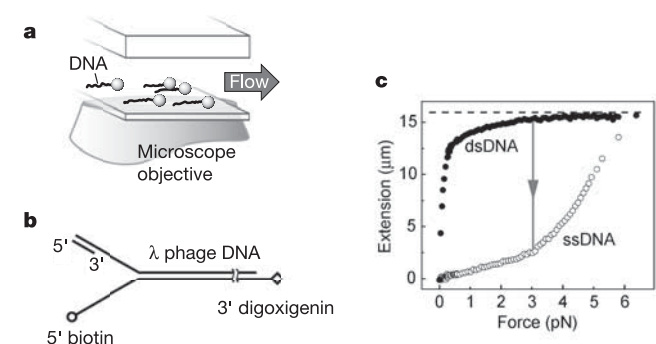


Figure 2 | Experimental design. **a**, Schematic representation of the experimental arrangement (not drawn to scale). **b**, Pre-made forked DNA substrate with single-stranded regions exposed at both 3' and 5' ends to allow for assembly of T7 DNA polymerase and gp4, respectively. **c**, Extension of λ -phage ssDNA (open circles) and dsDNA (filled circles) as a function of the applied stretching force. The dashed horizontal line at 16.2 μm represents the crystallographic length of B-form λ -phage dsDNA. The vertical arrow illustrates the change in length as a result of enzymatic conversion from dsDNA to ssDNA.

The coiling of single-stranded (ss)DNA causes it to be shorter than double-stranded (ds)DNA at low stretching forces (<6 pN; Fig. 2c). Consequently, conversion from dsDNA to ssDNA can be monitored through a decrease in total DNA length^{14,16,17}.

We preassemble a complex consisting of one T7 DNA polymerase (gp5-trx) and helicase/primase (gp4) on a forked λ phage DNA. This primer-template has an exposed 5' ssDNA tail to facilitate loading of the hexameric helicase¹⁸ and a DNA primer annealed to the other strand to initiate leading-strand synthesis (Fig. 2b). In the presence of deoxynucleoside 5' triphosphates (dNTPs), but in the absence of Mg^{2+} ions, only the gp4 and leading-strand DNA polymerase will remain stably bound to the DNA fork, without initiating synthesis¹⁹. Stringent washing of the flow cell after pre-assembly of the leading-strand complex removes all proteins not stably bound, including the lagging-strand DNA polymerase. Unwinding and nucleotide incorporation are started by addition of Mg^{2+} .

Leading-strand synthesis catalysed by T7 DNA polymerase converts one DNA strand arising from gp4 helicase activity into dsDNA²⁰. In the absence of the lagging-strand DNA polymerase, the lagging strand will remain in the single-stranded form. By attaching the DNA to the surface of the flow cell by the 5' biotin-labelled tail, leading-strand synthesis can be detected by an effective shortening of the DNA (Fig. 3a, b). The number of nucleotides polymerized can be obtained by using the difference between the lengths of ssDNA and dsDNA^{14,16,17}.

Averaged over multiple single-molecule traces ($n = 13$ traces), we measure the processivity of T7 leading-strand synthesis to be 17 ± 3 kb (mean $\pm$ s.e.m.), a dramatic increase compared to that of unwinding by the gp4 alone (60 base pairs (bp))²¹ or synthesis by the T7 DNA polymerase alone (~ 1 kb, see Supplementary Figs S1, S2 and ref. 10). The origin of this high processivity is likely to be found in a stabilization of the complex by a specific interaction of the acidic C-terminal domain of gp4 with basic regions on the T7 DNA polymerase¹⁹. The rate of leading-strand synthesis (164 ± 20 bp s^{-1} , $n = 13$) is identical to that measured in ensemble experiments²², confirming that the force exerted on the DNA has no influence on the enzymatic activities.

We enabled the primase activity of gp4 by adding the required ribonucleotides to the reaction mixture. The sequence recognized by the T7 gp4 primase is 3'CTG(G/T)(G/T)5', where the 3'C is essential for recognition but is not copied into the RNA primer²³. Consequently, the primase requires only the ribonucleotides rATP and rCTP. In the presence of both rATP and rCTP, we observe the appearance of short pauses in the single-molecule leading-strand synthesis traces (Fig. 3a, trace 2; pauses are indicated by grey arrows).

To confirm that the pauses originate from primase activity, we measured leading-strand synthesis catalysed by the T7 DNA polymerase and a variant of gp4 that is lacking the 63 N-terminal amino acids that form the zinc-binding domain (ZBD). This truncated gp4 cannot catalyse template-directed RNA synthesis²⁴, although it does form hexamers and show normal helicase activity²⁵. We performed leading-strand synthesis with this gp4 isoform both in the presence and absence of rATP and rCTP. The absence of pausing in either case confirms that the pausing is dependent on primase activity (Fig. 3a, traces 3 and 4). Pausing by wild-type gp4 is abolished when only a

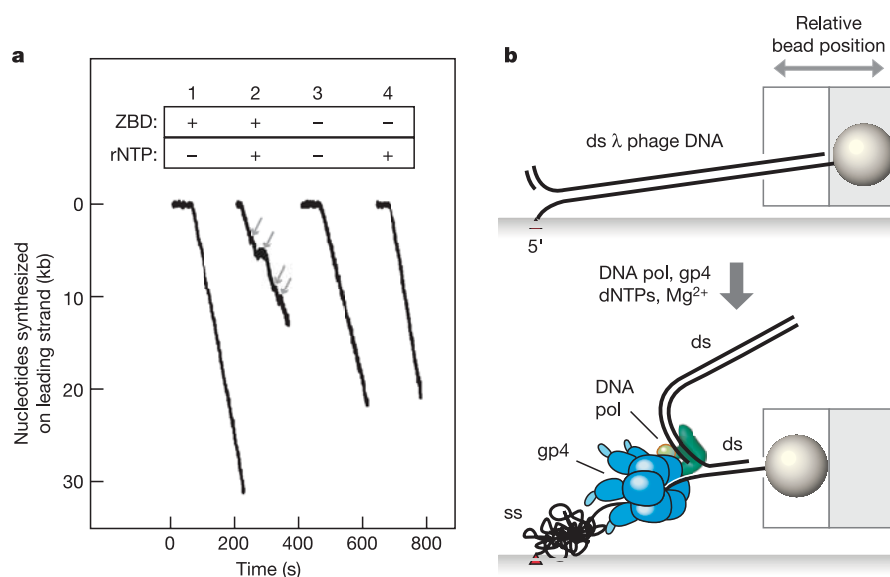


Figure 3 | Single-molecule observation of T7 leading-strand synthesis. **a**, Single-molecule trajectories. Above the traces is indicated whether the gp4 contained the zinc-binding domain (ZBD), and whether rATP and rCTP were present in the reaction mixture. Grey arrows denote pauses. **b**, Schematic depiction of events observed in **a**. In the absence of lagging-

strand synthesis, leading-strand synthesis causes the 5' tail of the DNA to be converted to the single-stranded form. Attachment of the 5' end to the surface allows us to monitor this conversion as a change in total length of the DNA.

single ribonucleotide is present (Supplementary Information). Furthermore, the pause sites in the presence of rATP and rCTP correspond to the positions of known primase recognition sites in the λ phage genome (Supplementary Information).

To exclude the possibility that pauses in leading-strand synthesis are simply caused by an inhibition of the primase to transfer its primer, we repeated the single-molecule replication experiments in the continuous presence of excess T7 DNA polymerase. The ensuing lagging-strand synthesis leads to the formation and release of a replication loop on the lagging strand, expressed in the single-molecule traces as a gradual shortening of the DNA followed by a prompt lengthening (Fig. 4). A quantitative analysis of the various length changes is presented in the Supplementary Information. The presence of replication loops confirms that correctly functioning replication forks are formed in our replication experiments. Loop formation is not observed when the ribonucleotides are withheld from the reaction. The low concentrations of DNA used in the single-molecule experiment result in a DNA–protein stoichiometry different from previously published bulk experiments⁸, and may explain the reduced processivity of the replisome.

The pauses before the initiation of replication loop formation represent a transient halting of the whole replication fork during primer synthesis and delivery. The average pause duration of 6 ± 2 s ($n = 5$) is similar to that measured in the leading-strand synthesis

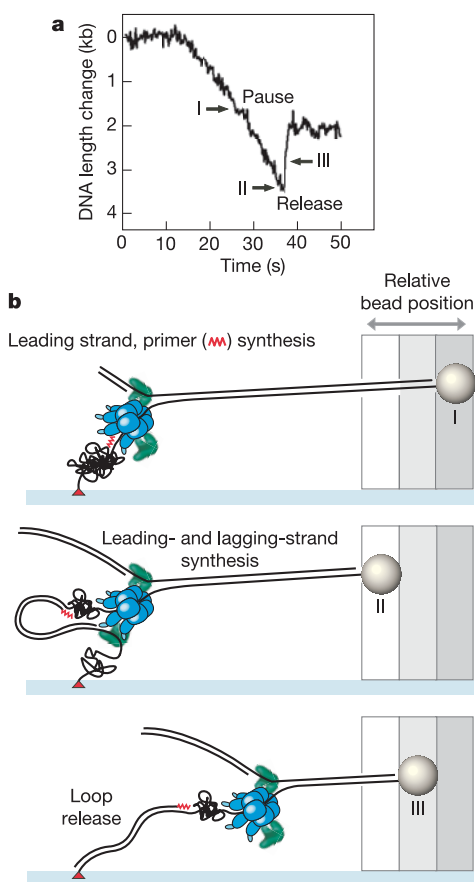


Figure 4 | Lagging-strand synthesis and loop formation. **a**, Single-molecule trajectory of replication loop formation. In the presence of additional T7 DNA polymerase, lagging-strand synthesis is initiated after primer synthesis (indicated by the pause), and a replication loop is formed in the lagging strand. Replication loop release is clearly visible as instantaneous lengthening of DNA. A quantitative interpretation of the different rates and length changes is presented in the Supplementary Information. **b**, Schematic depiction of the events observed in **a**.

experiments (5.6 ± 0.7 s, $n = 62$ at 3 pN; 5 ± 2 s, $n = 8$ at 1 pN), indicating that it is primarily determined by primer synthesis as opposed to delivery. This observation also confirms that the absence of the T7 single-stranded DNA-binding protein, gp2.5, is not likely to have an influence on the pausing behaviour of the fork. After all, gp2.5 mediates primer site recognition, but does not influence the RNA polymerization rate²⁶.

The several seconds needed to synthesize a tetraribonucleotide correspond well with the low rate of RNA synthesis by the primase as measured in ensemble experiments²³. In the time required by the primase to synthesize a primer, the leading-strand synthesis would have synthesized close to 1,000 bases. The fact that an Okazaki fragment in T7 replication has an average length of 3,000 nucleotides⁸ would mean that the lagging-strand DNA polymerase has to synthesize appreciably faster than the leading-strand polymerase to make up for this delay. However, closer analysis of replication loop formation as observed in our experiments shows that the rate of synthesis of the two DNA polymerases is equal (Supplementary Information). An alternative explanation could be that a primer for the next Okazaki fragment is already synthesized during the synthesis of the nascent fragment⁷, although the delay arising from the transfer of the lagging-strand DNA polymerase to the new primer remains. Our observation that leading-strand synthesis momentarily stops during primer synthesis provides a simpler solution. A transient halting of the whole replication complex during the slow enzymatic events on the lagging strand would allow the necessary transactions to be completed without leading-strand synthesis progressing too far ahead of the lagging-strand synthesis.

At this point, it is unclear what the molecular mechanism is that underlies the halting of leading-strand synthesis by the primase activity. The low affinity of the primase for the primer–template²³ makes it unlikely that stalling is caused by a direct primase–RNA–DNA interaction. This leaves the possibility that ribonucleotide condensation by the primase results in an allosteric regulation of helicase activity through protein–protein interactions. The replication machinery of other organisms, such as T4 and *E. coli*, differs from that of the T7 bacteriophage in that the primase is not covalently linked to the helicase. Nonetheless, oligomerization of the primase, with a similar organization as that observed for the T7 gp4, is achieved in these systems through the interaction between the primase and helicase, and has been demonstrated to be critical for primase synthesis^{27,28}. Albeit noncovalent in nature, this interaction could facilitate a similar regulation of helicase activity to that we have observed in the T7 replication machinery.

METHODS

Tethering and observation of single DNA molecules. Forked DNA molecules (see Supplementary Information for DNA substrate preparation) are attached by one end to the glass surface of a flow cell and by the other end to a bead (see ref. 14 and Supplementary Information). To prevent nonspecific interactions between the beads and the surface, we used paramagnetic beads with a diameter of 2.8 μ m (Dynal) and applied a small (1 pN) magnetic force upwards by positioning a permanent magnet above the flow cell. Combined with the horizontal drag force created by laminar flow, this magnetic force held the centre of the beads $\sim 4 \mu$ m above the surface of the flow cell. The flow cells were placed on an inverted microscope and the beads were imaged using a digital CCD camera. Typically, an area of 850μ m $\times$ 850μ m was imaged with a time-resolution of 200 ms. Particle-tracking software (Semasoph) was then used to determine the centre position of every bead for every frame with a precision of 1 nm (ref. 29). About 50 individual, tethered DNA molecules can be observed simultaneously, allowing for a high data throughput.

Replication reactions. T7 gp4 with and without the ZBD, and T7 gp5–trx were purified as previously described^{10,30}. The flow cell containing tethered, forked DNA molecules was incubated for 15 min with 18 nM T7 gp4 (hexameric) and 18 nM T7 gp5 (DNA polymerase) complexed 1:1 with *E. coli* thioredoxin in replication buffer (40 mM Tris pH 7.5, 50 mM potassium glutamate, 10 mM dithiothreitol, 100 μ g ml⁻¹ BSA), in the presence of 700 μ M each of dATP, dTTP, dCTP and dGTP. After washing with 15 flow cell volumes (300 μ l) of replication buffer and nucleotides, replication was started by introducing 10 mM MgCl₂,

700 μ M dNTPs and 300 μ M each of rATP and rCTP in replication buffer. Lagging-strand synthesis was initiated by adding 18 nM T7 gp5 to the latter mixture. All experiments were performed at a temperature of $22 \pm 1^\circ\text{C}$.

Data analysis. After particle tracking, traces were corrected for residual instabilities in the flow by subtracting traces corresponding to tethers that were not enzymatically altered. Bead displacements were converted into numbers of nucleotides synthesized, using the known length difference between ssDNA and dsDNA at our experimental conditions (Fig. 2c). The resultant traces were smoothed and fitted with series of line segments to detect pauses and determine rates. See Supplementary Information for details regarding data analysis.

Received 14 June; accepted 13 October 2005.

1. Alberts, B. DNA replication and recombination. *Nature* **421**, 431–435 (2003).
2. Sheaff, R. J. & Kuchta, R. D. Mechanism of calf thymus DNA primase: slow initiation, rapid polymerization, and intelligent termination. *Biochemistry* **32**, 3027–3037 (1993).
3. Swart, J. R. & Griep, M. A. Primer synthesis kinetics by *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry* **34**, 16097–16106 (1995).
4. Frick, D. N., Kumar, S. & Richardson, C. C. Interaction of ribonucleoside triphosphates with the gene 4 primase of bacteriophage T7. *J. Biol. Chem.* **274**, 35899–35907 (1999).
5. Stukenberg, P. T., Turner, J. & O'Donnell, M. An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. *Cell* **78**, 877–887 (1994).
6. Salinas, F. & Benkovic, S. J. Characterization of bacteriophage T4-coordinated leading- and lagging-strand synthesis on a minicircle substrate. *Proc. Natl Acad. Sci. USA* **97**, 7196–7201 (2000).
7. Tougu, K. & Marians, K. J. The interaction between helicase and primase sets the replication fork clock. *J. Biol. Chem.* **271**, 21398–21405 (1996).
8. Lee, J., Chastain, P. D. II, Griffith, J. D. & Richardson, C. C. Lagging strand synthesis in coordinated DNA synthesis by bacteriophage T7 replication proteins. *J. Mol. Biol.* **316**, 19–34 (2002).
9. Benkovic, S. J., Valentine, A. M. & Salinas, F. Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70**, 181–208 (2001).
10. Tabor, S., Huber, H. E. & Richardson, C. C. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262**, 16212–16223 (1987).
11. Guo, S., Tabor, S. & Richardson, C. C. The linker region between the helicase and primase domains of the bacteriophage T7 gene 4 protein is critical for hexamer formation. *J. Biol. Chem.* **274**, 30303–30309 (1999).
12. Frick, D. N., Baradaran, K. & Richardson, C. C. An N-terminal fragment of the gene 4 helicase/primase of bacteriophage T7 retains primase activity in the absence of helicase activity. *Proc. Natl Acad. Sci. USA* **95**, 7957–7962 (1998).
13. Kusakabe, T., Baradaran, K., Lee, J. & Richardson, C. C. Roles of the helicase and primase domain of the gene 4 protein of bacteriophage T7 in accessing the primase recognition site. *EMBO J.* **17**, 1542–1552 (1998).
14. van Oijen, A. M. *et al.* Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235–1238 (2003).
15. Bustamante, C., Smith, S. B., Liphardt, J. & Smith, D. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* **10**, 279–285 (2000).
16. Wuite, G. J., Smith, S. B., Young, M., Keller, D. & Bustamante, C. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* **404**, 103–106 (2000).
17. Maier, B., Bensimon, D. & Croquette, V. Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc. Natl Acad. Sci. USA* **97**, 12002–12007 (2000).
18. Ahnert, P., Picha, K. M. & Patel, S. S. A ring-opening mechanism for DNA binding in the central channel of the T7 helicase-primase protein. *EMBO J.* **19**, 3418–3427 (2000).
19. Hamdan, S. M. *et al.* A unique loop in T7 DNA polymerase mediates the binding of helicase-primase, DNA binding protein, and processivity factor. *Proc. Natl Acad. Sci. USA* **102**, 5096–5101 (2005).
20. Kolodner, R. & Richardson, C. C. Gene 4 protein of bacteriophage T7. Characterization of the product synthesized by the T7 DNA polymerase and gene 4 protein in the absence of ribonucleoside 5'-triphosphates. *J. Biol. Chem.* **253**, 574–584 (1978).
21. Jeong, Y. J., Levin, M. K. & Patel, S. S. The DNA-unwinding mechanism of the ring helicase of bacteriophage T7. *Proc. Natl Acad. Sci. USA* **101**, 7264–7269 (2004).
22. Stano, N. M. *et al.* DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature* **435**, 370–373 (2005).
23. Frick, D. N. & Richardson, C. C. Interaction of bacteriophage T7 gene 4 primase with its template recognition site. *J. Biol. Chem.* **274**, 35889–35898 (1999).
24. Bernstein, J. A. & Richardson, C. C. A 7-kDa region of the bacteriophage T7 gene 4 protein is required for primase but not for helicase activity. *Proc. Natl Acad. Sci. USA* **85**, 396–400 (1988).
25. Bernstein, J. A. & Richardson, C. C. Purification of the 56-kDa component of the bacteriophage T7 primase/helicase and characterization of its nucleoside 5'-triphosphatase activity. *J. Biol. Chem.* **263**, 14891–14899 (1988).
26. He, Z.-G. & Richardson, C. C. Effect of single-stranded DNA-binding proteins on the helicase and primase activities of the bacteriophage T7 gene 4 protein. *J. Biol. Chem.* **279**, 22190–22197 (2004).
27. Yang, J., Xi, J., Zhuang, Z. & Benkovic, S. J. The oligomeric T4 primase is the functional form during replication. *J. Biol. Chem.* **280**, 25416–25423 (2005).
28. Mitkova, A. V., Khopde, S. M. & Biswas, S. B. Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of *Escherichia coli* in RNA primer synthesis. *J. Biol. Chem.* **278**, 52253–52261 (2005).
29. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).
30. Notarnicola, S. M., Mulcahy, H. L., Lee, J. & Richardson, C. C. The acidic carboxyl terminus of the bacteriophage T7 gene 4 helicase/primase interacts with T7 DNA polymerase. *J. Biol. Chem.* **272**, 18425–18433 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We wish to thank T. Ellenberger, D. Crampton and P. Blainey for discussions and comments, and S. Buratowski for critically reading the manuscript. We are grateful to S.-J. Lee for providing the purified ZBD-less gp4 variant. We thank S. Moskowitz for preparation of the figures. This work was supported by grants from the NIH to C.C.R. and X.S.X.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M.v.O. (antoine_van_oijen@hms.harvard.edu).

Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state

Holger Rehmann¹, Joost Das¹, Puck Knipscheer², Alfred Wittinghofer³ & Johannes L. Bos¹

Epac proteins (exchange proteins directly activated by cAMP) are guanine-nucleotide-exchange factors (GEFs) for the small GTP-binding proteins Rap1 and Rap2 that are directly regulated by the second messenger cyclic AMP^{1,2} and function in the control of diverse cellular processes, including cell adhesion and insulin secretion³. Here we report the three-dimensional structure of full-length Epac2, a 110-kDa protein that contains an amino-terminal regulatory region with two cyclic-nucleotide-binding domains and a carboxy-terminal catalytic region. The structure was solved in the absence of cAMP and shows the auto-inhibited state of Epac. The regulatory region is positioned with respect to the catalytic region by a rigid, tripartite β -sheet-like structure we refer to as the 'switchboard' and an ionic interaction we call the 'ionic latch'. As a consequence of this arrangement, the access of Rap to the catalytic site is sterically blocked. Mutational analysis suggests a model for cAMP-induced Epac activation with rigid body movement of the regulatory region, the features of which are universally conserved in cAMP-regulated proteins.

The N-terminal regulatory region of Epac (Fig. 1a) is an auto-inhibitory domain that blocks GEF activity for Rap proteins. The second cyclic-nucleotide-binding domain (cNBD-B) of Epac2, which is common to Epac1 and Epac2, is sufficient to keep Epac inactive⁴. The first cNBD of Epac2 (cNBD-A) binds cAMP with relatively low affinity⁴, and its function is currently unclear. The regulatory region also contains a Dishevelled, Egl-10, Pleckstrin (DEP) domain, which is involved in membrane localization³. The C-terminal catalytic region comprises a Ras-exchange motif (REM) domain, a Ras-association domain and a CDC25-homology domain (CDC25-HD). To understand how Epac is regulated, we have determined the crystal structure of full-length Epac2 in the absence of cAMP, to a resolution of 2.7 Å (Fig. 1b, c and Supplementary Table 1).

A ribbon diagram of the protein shows a very compact arrangement of the domains in the regulatory and catalytic regions. In spite of the insertion of the Ras-association domain in the catalytic region, the structures and relative orientation of the CDC25-HD and REM domains are very similar to those found for a catalytic fragment of the Ras-GEF Sos in complex with nucleotide-free Ras⁵ (Supplementary Fig. 1). The binding site of Rap on Epac is thus unambiguous. A distal binding site for RasGTP is located in the REM domain of Sos, serving as a positive-feedback loop⁶. This putative distal binding site is accessible in Epac but only partially conserved (Supplementary Fig. 2).

A characteristic feature of the CDC25-HD domain is a helical hairpin (HP in figures) formed by two antiparallel α -helices pointing outwards (Fig. 1b). The C-terminal helix of this hairpin (lower dark blue helix in Fig. 1) interacts with the REM domain, whereas the N-terminal helix (upper dark blue helix in Fig. 1) is a crucial

part of the binding site for the nucleotide-free GTP-binding protein (Fig. 1c, d). The helical hairpin displaces the switch 1 region of the GTP-binding protein and thereby catalyses nucleotide release⁵.

The extra domain in Epac between the REM domain and the CDC25-HD domain was predicted to be a Ras-association domain (for example, in the Pfam database), and indeed, this domain adopts a ubiquitin fold, which is characteristic of Ras-association domain⁷. However, unlike ubiquitin and the Ras-association domains of RalGDS⁹, the β -strands (as analysed by DSSP¹⁰) are only weakly connected by main-chain interactions (Supplementary Fig. 3). Although the function of this domain in Epac2 is still elusive, a similar domain in the RapGEFs Rapac and PDZ-GEF interacts with members of the Ras family^{11,12}.

Superposition of the Ras-Sos complex onto Epac results in a number of serious steric clashes of Ras, mainly with the first but also with the second cNBD (Fig. 1d). By extension, the position of the regulatory region excludes binding of Rap to Epac2 by steric hindrance. By removing the steric block by the regulatory region, the CDC25-HD should be able to accept Rap binding and induce nucleotide release, as superimposition of Ras-bound Sos with the catalytic region of Epac2 shows only minor deviations (Supplementary Fig. 1). Likewise, the structure of the regulatory region, with the exception of some minor differences in loop regions, is identical to the structure of the isolated regulatory region determined previously¹³ (Supplementary Fig. 4). cNBD-A and cNBD-B are facing each other, so that a movement of cNBD-A is required to allow access of cAMP to the binding sites. Indeed, the linker between cNBD-A and the DEP domain is flexible and not visible in the structures.

Activation of Epac thus requires movement of the regulatory region away from the catalytic region to allow Rap binding. This is facilitated by the loose connection between the two halves of Epac, which are secured by only two contact points (Fig. 1b, c). One is a five-strand, β -sheet-like tripartite structure formed by the C terminus of cNBD-B, the N terminus of the REM domain and the loop of the helical hairpin in CDC25-HD (Figs 1b, 2a), and constitutes the central 'switchboard' (SB in figures) of Epac. The second strand of this structure is formed by the residues VLVLE, a highly conserved sequence in the C-terminal end of Epac cNBDs from fruitfly to human. A mutant of Epac1 in which the VLVLE sequence is changed to polyalanine is constitutively active¹⁴, indicating that the integrity of the switchboard is essential for maintaining the relative orientation between the regulatory region and the catalytic region, and thus for keeping Epac inactive.

The loop of the catalytic helical hairpin is part of the β -sheet-like structure, and thus part of the contact points between the regulatory and catalytic regions. This part of the helical hairpin is not involved in Rap binding. As in the Ras-Sos complex (Supplementary Fig. 1), the second helix of the hairpin is contacted and stabilized by the REM

¹Department of Physiological Chemistry and Centre for Biomedical Genetics, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands.

²Division of Molecular Carcinogenesis and Centre for Biomedical Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

³Max-Planck-Institut für Molekulare Physiologie, Otto-Hahn-Straße 11, D-44227, Dortmund, Germany.

domain, with Phe 930 forming an extensive hydrophobic core with the REM domain (Figs 1b, 2b). The strand of the REM domain involved in the switchboard is sandwiched between the VLVLE strand of cNBD-B and the loop of the helical hairpin, which is much shorter than in Sos. The interactions provided by the loop of the helical hairpin are fewer and resemble a regular β -sheet arrangement to a lesser extent. The structure of the switchboard determines the relative orientation of the cNBD-B, the REM and the catalytic domains. On the basis of the large number of main-chain interactions, we assume that this regulatory element will not change its structure and that only the orientation of the cNBD-B relative to the switchboard will be modified after cAMP binding (see below).

The second contact point between the regulatory and catalytic regions is an ionic latch formed between Arg 886 and Asp 883 from CDC25-HD, and Gln 303, Asp 307 and Glu 332 of cNBD-B (Fig. 2c).

Gln 303 and Asp 307 are found in the long connecting α -helix between the DEP domain and cNBD-B. Notably, Glu 332 is located in a region of the N-terminal helical bundle of cNBD-B, which undergoes conformational rearrangements upon cAMP binding in cyclic-nucleotide-regulated ion channels¹⁵ and protein kinase A (PKA)^{16,17}. Arg 886 and Asp 883 are located at the periphery of the Ras-interaction surface, as inferred from the Ras–Sos complex⁵.

We analysed the function of this interaction in a mutagenesis study. In the Epac2 Δ 280 mutant, cNBD-A and the DEP domain are deleted, but the connecting α -helix with the interacting residues is still present. In Epac2 Δ 306, part of the connecting α -helix, including the interacting residues, is removed as well. The affinities of Epac2 Δ 280 and Epac2 Δ 306 for cAMP are almost identical (Fig. 3). The maximum exchange activity under saturating cAMP concentrations ($k_{\max}$) of Epac2 Δ 280 is very similar to that of the wild-type

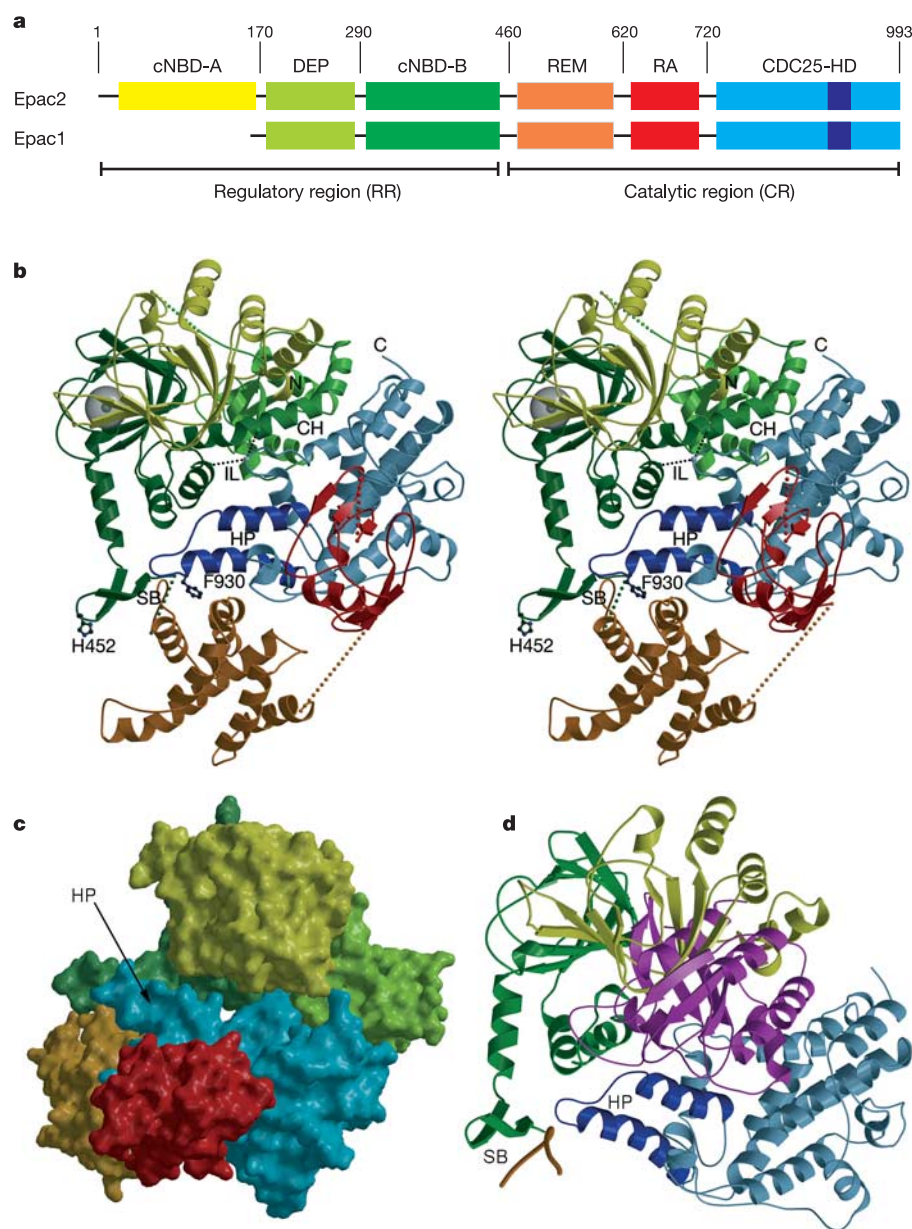


Figure 1 | Structure of Epac2. **a**, Domain organization of Epac. The same colour code is used throughout the figures. CDC25-HD, CDC25-homology domain; cNBD, cyclic-nucleotide-binding domain; DEP, Dishevelled, Egl-10, Pleckstrin domain; RA, Ras-association domain; REM, Ras-exchange motif. **b**, Ribbon diagram of Epac2 in stereo view. Missing connectivity is indicated by coloured dotted lines. The green ball indicates the cAMP-

binding site in cNBD-B. HP, helical hairpin of the CDC25-HD (dark blue); SB, switchboard; IL, ionic latch (dotted black lines); CH, connecting helix (helix H4 in ref 13). **c**, Surface representation of Epac.

d, Superposition of Epac and the Ras–Sos complex. The RA, DEP and REM domains are omitted. Only Ras (magenta) from the Ras–Sos complex is shown.

protein (data not shown). Importantly, a fivefold increase in $k_{\max}$ is observed for Epac2 Δ 306 compared with Epac2 Δ 280 (Fig. 3). This result is in agreement with the idea that even in the presence of cAMP, Epac exists in equilibrium between an inactive and active state^{14,18}. After disruption of the ionic latch, the inactive state is much less favoured, resulting in the increased $k_{\max}$ of Epac2 Δ 306.

We next generated the mutant Epac2 Δ 280 R886A. To our surprise, we found a dramatic reduction in GEF activity, which was still cAMP-dependent (Fig. 3). It would seem that Arg 886 has a crucial role in the interaction with Rap, indicating that the interaction between the regulatory and catalytic regions directly blocks at least one important residue for the interaction between Epac and Rap.

Binding of cAMP to the second cNBD results in the activation of Epac⁴. Based on the comparison of the cAMP-free regulatory region of Epac with cAMP-bound PKA, we have previously suggested that

binding of cAMP to a cNBD causes a reorientation of the phosphate-binding cassette (PBC)¹³. The PBC is highly conserved in cNBD domains and interacts with the phosphate and sugar moieties of the cyclic nucleotide¹⁹. The reorientation would induce a conformational change of conserved hydrophobic residues in the PBC, and allow the C-terminal helix of the cNBD—termed the 'hinge'—to move closer to the core of the domain (Fig. 4). This model was recently confirmed by the structures of cyclic-nucleotide-regulated ion channels with and without cAMP¹⁵, and a cAMP-free holoenzyme construct of PKA¹⁷. In the active state, the region C-terminal to the hinge forms a lid that binds to the base of cAMP and protects it from the solvent. In Epac this region corresponds to the first two strands of the switchboard.

Supporting the mechanism by which these β -strands would form the lid that contacts the base of the cyclic nucleotide, mutations of conserved residues in or flanking the first β -strand interfere with the activation of Epac1 by cAMP and cAMP analogues^{14,18}. Strikingly, an H317A mutation in Epac1 almost completely abolishes activation without interfering with cAMP-binding itself. Similarly, the H452A mutation reduces $k_{\max}$ in Epac2 (Fig. 3). His 452 is positioned in the loop between strands 1 and 2, completely exposed to the solvent and far away from the cAMP-binding site (Figs 1b, 2a). Similarly, mutations of Glu 308, Thr 311 and Arg 313, which are solvent-exposed in the structure of Epac2, interfere with communication

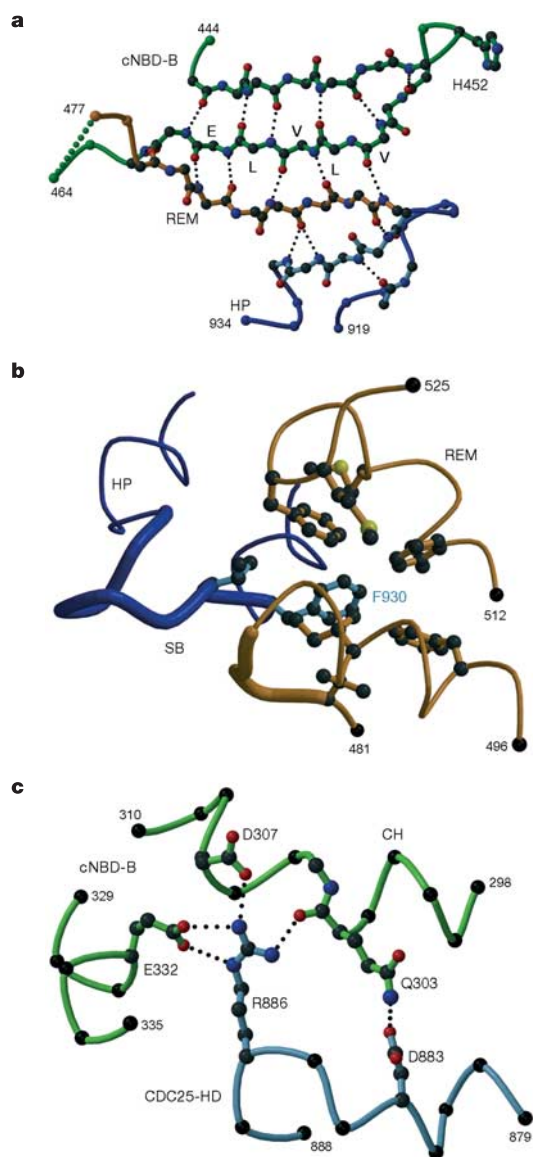


Figure 2 | Anchoring points between the regulatory and catalytic regions. **a**, The switchboard is formed by strands provided by cNBD-B (dark green), the REM domain (orange) and the loop of the helical hairpin (HP) of CDC25-HD (blue). Peptides are reduced to polyglycine. Hydrogen bonding by main-chain atoms is indicated by dotted lines. **b**, The REM domain (orange) interacts tightly with the C-terminal helix of the helical hairpin (blue). Thick C α traces highlight the parts of the switchboard shown in **a**. **c**, The ionic latch between cNBD-B (green) and CDC25-HD (blue). CH, connecting helix (see Fig. 1b).

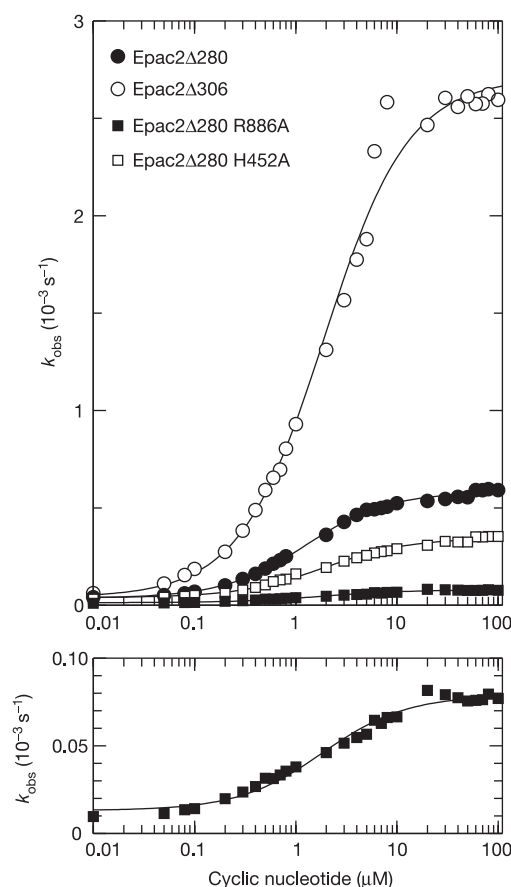


Figure 3 | Dependence of Rap-GEF activity (as k_{obs}) on cAMP concentration. Exchange activities for Epac2 Δ 280, Epac2 Δ 280 R886A, Epac2 Δ 280 H452A and Epac2 Δ 306 are shown. The lower panel shows a magnification of the results for Epac2 Δ 280 R886A. Increasing concentrations of cAMP were added to a mixture containing 200 nM Rap-mGDP and 300 nM Epac, and the exchange against 100-fold excess of unlabelled GDP was measured as time-dependent decay fluorescence, fitted to a single exponential to obtain k_{obs} . Half-maximal activities for Epac2 Δ 280, Epac2 Δ 280 H452A, Epac2 Δ 280 R886A and Epac2 Δ 306 are 1.1 μ M, 1.8 μ M, 1.7 μ M and 1.6 μ M, respectively.

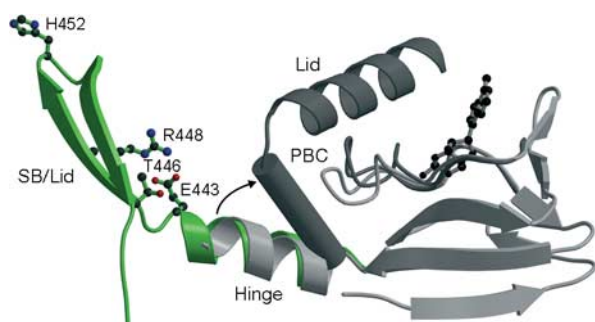


Figure 4 | Comparison of Epac with cyclic-nucleotide-regulated ion channels. cNBD-B of Epac and the cAMP-bound and cAMP-free cNBDs of ion channels are superimposed on one another. The cAMP-free channel is shown in light grey. Regions that differ in the cAMP-bound state are shown in dark grey, with cAMP shown in ball-and-stick representation. The cNBD-B of Epac is very similar to the cAMP-free channel. In Epac, the C-terminal helix, which functions as the hinge, and the β -strands, which are part of the switchboard (SB) and are assumed to form the lid, are shown in green. PBC, phosphate-binding cassette.

between cAMP binding and activation in Epac1^{14,18}. It is therefore highly likely that the β -strands at the C terminus of the cNBD-B domain serve as the lid to cover bound cAMP. Movement of the second cNBD towards this lid would uncover the Rap-binding site in CDC25-HD (Supplementary Fig. 5).

In the three classes of eukaryotic cAMP-regulated proteins, three different modes of inhibition are found. In PKA, the cNBD of the regulatory subunit interacts over an extensive surface area with the catalytic subunit to place a pseudo-substrate sequence in optimal proximity to the catalytic cleft of the kinase. cAMP binding interferes with the integrity of the interaction surface¹⁷. In cyclic-nucleotide-regulated channels, the N-terminal helices of the cNBD are assumed to be part of the ionic pore²⁰ and are shown to undergo conformational rearrangements upon cyclic nucleotide binding¹⁵. In Epac, the access of Rap to the catalytic site is sterically blocked. cAMP binding liberates cNBD-B, allowing it to leave its inhibitory position. In all three cases, the required conformational changes are based on the rearrangement of the PBC caused by cAMP, which allows the hinge to move closer to the core of the domain and allows the lid to adopt its cAMP-binding position^{13,15,17}.

METHODS

Protein preparation. Epac2 proteins (RapGEF4, *Mus musculus*, accession number AF115480) were expressed as glutathione S-transferase (GST)-fusion proteins (pGEX4T2, Pharmacia) in bacteria, similar to the procedure described for Epac1 (ref. 14).

Crystallography. Crystals of full-length Epac2 (amino acids 1–993) were grown at 293 K in sitting drop set-ups using a reservoir solution containing 100 mM Bis-Tris propane pH 7.5, 200 mM NaNO₃, 12% PEG 3350 and a protein concentration of 10 g l⁻¹. Crystals were flash-frozen in liquid nitrogen with a solution containing the mother liquor and 16% glycerol. Data were collected at 100 K. Crystals diffract to 2.7 Å resolution and belong to the orthorhombic space group *P*₂₁₂₁ with one molecule per asymmetric unit. For further details see Supplementary Methods.

Biochemistry. GEF assays were performed as described previously¹⁸. In brief, Rap was loaded with the fluorescent GDP analogue 2'-/3'-O-(N'-methyl-anthraniloyl)-guanosinediphosphate (mGDP) (Biolog) and incubated with an excess of unlabelled GDP for measurements. The fluorescence intensity of

Rap-bound mGDP is approximately twice as intense as free mGDP and thus nucleotide exchange can be observed as a decay in fluorescence. The speed of decay depends on Epac activity and thus on cAMP concentration.

Received 25 October; accepted 22 November 2005.

- de Rooij, J. *et al.* Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* **396**, 474–477 (1998).
- Kawasaki, H. *et al.* A family of cAMP-binding proteins that directly activate Rap1. *Science* **282**, 2275–2279 (1998).
- Bos, J. L. Epac: a new cAMP target and new avenues in cAMP research. *Nature Rev. Mol. Cell Biol.* **4**, 733–738 (2003).
- de Rooij, J. *et al.* Mechanism of regulation of the Epac family of cAMP-dependent RapGEFs. *J. Biol. Chem.* **275**, 20829–20836 (2000).
- Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337–343 (1998).
- Margarit, S. M. *et al.* Structural evidence for feedback activation by Ras-GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* **112**, 685–695 (2003).
- Herrmann, C. Ras-effector interactions: after one decade. *Curr. Opin. Struct. Biol.* **13**, 122–129 (2003).
- Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).
- Huang, L., Hofer, F., Martin, G. S. & Kim, S. H. Structural basis for the interaction of Ras with RalGDS. *Nature Struct. Biol.* **5**, 422–426 (1998).
- Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
- Liao, Y. *et al.* RA-GEF, a novel Rap1A guanine nucleotide exchange factor containing a Ras/Rap1A-associating domain, is conserved between nematode and humans. *J. Biol. Chem.* **274**, 37815–37820 (1999).
- Rebhun, J. F., Castro, A. F. & Quilliam, L. A. Identification of guanine nucleotide exchange factors (GEFs) for the Rap1 GTPase. Regulation of MR-GEF by M-Ras-GTP interaction. *J. Biol. Chem.* **275**, 34901–34908 (2000).
- Rehmann, H. *et al.* Structure and regulation of the cAMP-binding domains of Epac2. *Nature Struct. Biol.* **10**, 26–32 (2003).
- Rehmann, H., Rueppel, A., Bos, J. L. & Wittinghofer, A. Communication between the regulatory and the catalytic region of the cAMP-responsive guanine nucleotide exchange factor Epac. *J. Biol. Chem.* **278**, 23508–23514 (2003).
- Clayton, G. M., Silverman, W. R., Heginbotham, L. & Morais-Cabral, J. H. Structural basis of ligand activation in a cyclic nucleotide regulated potassium channel. *Cell* **119**, 615–627 (2004).
- Su, Y. *et al.* Regulatory subunit of protein kinase A: structure of deletion mutant with cAMP binding domains. *Science* **269**, 807–813 (1995).
- Kim, C., Xuong, N. H. & Taylor, S. S. Crystal structure of a complex between the catalytic and regulatory (R α) subunits of PKA. *Science* **307**, 690–696 (2005).
- Rehmann, H., Schwede, F., Doskeland, S. O., Wittinghofer, A. & Bos, J. L. Ligand-mediated activation of the cAMP-responsive guanine nucleotide exchange factor Epac. *J. Biol. Chem.* **278**, 38548–38556 (2003).
- Canaves, J. M. & Taylor, S. S. Classification and phylogenetic analysis of the cAMP-dependent protein kinase regulatory subunit family. *J. Mol. Evol.* **54**, 17–19 (2002).
- Zagotta, W. N. *et al.* Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425**, 200–205 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Sixma for access to crystallization robots. We thank M. Weyand and A. Scrima for data collection and discussions. H.R. was supported by the Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW) and is a recipient of the Otto-Hahn-Medaille of the Max-Planck-Gesellschaft. J.D. was supported by the Dutch Cancer Society.

Author Information Structure coordinates for Epac2 have been deposited in the Protein Data Bank under accession number 2BYV. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for material should be addressed to J.L.B. (J.L.Bos@med.uu.nl).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Stress management

Pressure; life as a student is full of it. It can lead to stress, depression, loss of sleep and, in some cases, suicide. At a meeting sponsored by scientific research society Sigma Xi and the US National Postdoctoral Association (NPA) last month, a few university administrators talked about their approaches to the problem.

A few years ago, the University of California, Berkeley, was shocked when three postdocs committed suicide within 18 months of each other. The event highlighted flaws in the support given to students: only one of the three was considered an employee, and so would have been eligible to use the university's mental-health facilities. And even then, it became clear that these facilities weren't advertised particularly well to postdocs. That has now changed, says Sam Castañeda, the university's director of postdoc affairs. The university last year modified its health system to provide access to all postdocs at all of its campuses, regardless of their employment status.

In another approach, the University of Vanderbilt School of Medicine in Nashville, Tennessee, recently hired a clinical psychologist to work out of the university's postdoc office.

Roger Chalkley, senior associate dean of the the medical school, says that research is intense and can give rise to myriad pressures whose manifestations "are best handled by a professional". Access to the psychologist will not be dictated by employment status and is available to postdocs, graduate students and medical students.

Could such resources become more widespread in the United States? Alyson Reed, executive director of the NPA, says for that to happen, institutions must either follow the University of California's lead and open up access to all postdocs, regardless of funding, or else the National Institutes of Health must change its designation of postdocs, as under many grants they can't be considered employees. Either approach would ease the pressure and make research training a safer, healthier proposition.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

MOVERS

Ernst Ulrich von Weizsäcker, dean, Donald Bren School of Environmental Science and Management, University of California, Santa Barbara



1998-2005: Member of the Bundestag, Berlin, Germany
1991-2000: Founding president, Wuppertal Institute for Climate, Environment and Energy, Berlin, Germany
1984-91: Director, Institute for European Environmental Policy, Bonn, Germany

Ernst Ulrich von Weizsäcker is proof that scientific disciplines don't define destiny — personal energy does. He has worked on projects on a wide range of topics, from climate and energy to world peace, as both a researcher and a politician. This isn't too surprising: von Weizsäcker studied chemistry and physics at the University of Hamburg in Germany, only to switch to biology for a PhD on honeybee vision at Freiburg University. "I always sensed that the mainstream disciplinary approach was exhausted — that the new fields of interests were located at the interface of existing disciplines," he says.

Always an environmentalist, who also describes himself as an opportunist, he has sought a highly varied career. After finishing graduate school during the cold war, he used his natural-sciences background to do research on peace at the Protestant Research Institute in Heidelberg, Germany. Following both a professorship in interdisciplinary biology at the University of Essen and his involvement with the German Social Democratic Party, he wrote a book on university reform. This interest inspired him not only to become founding president of the University of Kassel, but to create several interdisciplinary research centres, including the first in Germany devoted to organic farming.

In 1991, he formed the Wuppertal Institute for Climate, Environment and Energy, a think-tank seeking radical policy approaches to eco-efficiency: maximizing the value of products or services while reducing their environmental burden. He realized that the obstacles had less to do with scientific knowledge and more to do with policy.

He was so passionate about the eco-efficiency cause that he successfully ran for a seat in Germany's parliament, the Bundestag. For the past three years, he was chairman of the Bundestag Environment Committee, which pushed to align German environmental policies with European ones. "That was not overly successful," he admits. Politics is quite a difficult thing compared with the sciences, he says.

Resisting invitations to run again for parliament, he accepted an offer to direct the Bren School of Environmental Science and Management at the University of California, Santa Barbara. He was impressed by its interdisciplinary structure, which combines science with management.

Given his varied career, von Weizsäcker's advice is not surprising. "Be bold. Don't stick to just one discipline. Use your curiosity in another discipline — and ask the questions those trained in the field would not ask."

Virginia Gewin

RECRUITERS & INDUSTRY

Grooming women entrepreneurs

Starting your own company is risky, and even more so when it is a life-science company. More and more women scientists are starting their own biotech companies, but they face many challenges in their transition to the business world. They often lack the business skills and networking savvy to obtain the resources needed to move their research from concept to product. Women In Bio (WIB) was formed to help budding entrepreneurs move from the bench to the board room and to help women advance in their life-science careers.

WIB offers networking opportunities, access to resources essential for business growth, and educational programmes on two tracks: corporate and professional development. On the corporate development programme, for example, our Entrepreneur and Executive Forum provides an intimate environment to meet with experts and have frank, open discussions on selected business topics, such as the role of the US Food and Drug Administration in drug development, potential liability for the life-sciences entrepreneur, and strategies for licensing of intellectual property.

The professional development programmes are aimed at building skills that may have an important impact on career advancement and career

satisfaction. Workshops are open to women in all fields of life science and centre on enhancing communication and negotiation skills. The goal is to equip women to feel more confident about expressing themselves in their professional and personal lives. A recent workshop, "Speaking Your Mind and Getting Heard", dealt with how to get one's ideas heard in meetings.

Networking is critical in career advancement and business growth. Networking events, such as an annual dinner, open up opportunities to meet women on different career paths and further explore the business world.

WIB members are entrepreneurs, executives, scientists, investors and professionals from life-sciences companies, academia, government and professional organizations across the United States. Most members live in Maryland and Washington DC and about a third are scientists, many at federal labs such as the National Institutes of Health, and include students and postdoctoral researchers.

One of our members, Mona Jhaveri, a cancer researcher, is starting up her own company to develop cancer drugs. She credits WIB for connecting her to other women who are now guiding her in her new venture.

Robbie Melton is the president and director of Women in Bio.

ALUMNUS JOURNAL

A new chapter

The year since I was a *Naturejobs* Graduate Journal writer was one of the most eventful times of my life. In my last journal entry, I wondered how I could move from research into a 'bigger picture' job, where I could share the excitement of science with a wider audience. As it turned out, my ramblings in *Naturejobs* had more impact than I imagined. Through my writing, I met people at *Nature*. Last May, while I was still working in the lab, I received an e-mail encouraging me to apply for a position that had opened up. I jumped at the chance and landed myself a dream job: for nine months I was going to be one of *Nature's* manuscript editors.

Juggling both thesis and job is not for the faint-hearted, but I would do it all over again. Working at *Nature* has been everything I wished for and more. Every day, cutting-edge papers land on my desk. I interact with some of the world's best scientists. I am an editor and writer, yet still in the thick of research. It's a massive responsibility, but an immense privilege.

Now a new opportunity has arisen. Next year, I will be in San Francisco, helping to launch a new journal under the *Nature* umbrella. I can't wait. And in the midst of all this, I found the love of my life. San Francisco will be a new chapter for both of us. They say that life is full of surprises, and they are right. My advice to graduate students looking for direction: act on crazy ideas, never give in, and go with the flow. Something will come up.

Amber Jenkins is an associate editor with Nature Publishing Group.

Hiking the roof of the world

Just a few simple preparations.

Richard A. Lovett

Courtney Brandt was having a nightmare when the music started. One moment she was floundering in a blizzard — lost, gasping for air that wasn't there: all the things that most emphatically were not going to happen. The next, she was hearing Pachelbel's *Canon*, which for some reason she'd thought right for this, her first climb of a peak higher than Ben Nevis. Silly idea. Pachelbel wasn't Everest music. Even sillier had been reading a history chip of the days before smart fabrics and solar nets — the type of stuff that makes you think too much about the environment against which your equipment is designed to shield you.

Outside, that environment was making its own music: a symphony of moans, shrieks and wails as wind whipped through the South Col. But inside, the smart-tent was damping it nicely. The rental agent in Kathmandu had said the walls formed a 360 × 360 speaker web for reverse-phase damping, but however it worked, it reduced the outer world to a comforting murmur, like surf at her uncle's beach house.

She stirred and the music died. So did the background damping, allowing her to fully appreciate the gale raging outside. It was hard to believe that in an hour it would be flat calm. But that was what the weather-control folks promised, and they always delivered.

Not that Everest hikers had the clout to order up weather, despite the extortionate permit fees. That was a by-product of monsoon control. Over in Bangladesh, farmers got bumper crops of rice. Here, hikers snagged six perfect days, although Courtney wished she'd not drawn a start time in the first wave. Rather than hiking to the col in sun, like the lucky folks still in the Base Camp car park, she'd had to ascend in weather nearly as bad as Scotland at its nastiest. Oh well, you didn't turn down Everest just because you got a crappy start time.

Already, she could hear the wind dying. She wanted nothing more than to get started, but the tent wasn't going to let her out until it ran her through an extended checklist, now displayed on a section of touch-fabric.

She acknowledged the first few panels

with quick finger jabs. The batteries were fine. She'd brought her own solar net just to be on the safe side, but she was also plugged into the web of left-behind nets sprawling across half the col. Even with yesterday's clouds and two-dozen other parties using them, there'd been wattage to burn.

Next came the diagnostics for her thermal underwear. Dutifully, she checked the batteries — not that she'd be likely to need them. So long as she kept moving, the piezoelectric fabric would draw enough power from her body motion for the demon-pores to do their thing. That odd name, if she remembered correctly, had something to do with a guy named Maxwell. Powered by the combo of piezoelectrics, batteries and a solar-web knitted into her parka, the Maxwell filters would keep her warm by separating 'hot' air molecules from 'cold' ones. Turn them up too high, and you could cook yourself in Antarctica.

Finally came the oxygen web. More selective permeability, only this time chemical, not thermal. Oxygen in, everything else out. It made her look like a jungle explorer with a hat-brim of mosquito netting.

Unlike the tent, the O₂ web wasn't pressurized. But with the oxygen confined to the vicinity of her face, there was no fire danger and you could dial it up to 100% if you wanted, which was undoubtedly what today would call for. As a backup, a tube ran beneath her nose, ready to feed oxygen

from a membrane in her scarf if the jungle-hat set-up failed. Her equipment had so many redundancies, in fact, that she wanted to yell at the tent to just forget them and let her go.

At last, when she'd swallowed the pills that increased her blood's affinity for oxygen, the tent agreed she was ready. A door formed from an invisible seam, the over-pressure souged out, and she slipped through behind it. Nearby, other climbers were emerging from their tents. Some were in pairs or trios, but most, like Courtney, were solo. She pushed the stud that extruded the crampons from her chemically heated boots, clipped onto the monofilament rope that led all the way to the summit, dialled her oxygen to max, and started to climb. She was in good shape, and with plenty of oxygen, no worries about getting cold and no way to fall off, Everest wasn't really all that much tougher than Ben Nevis.

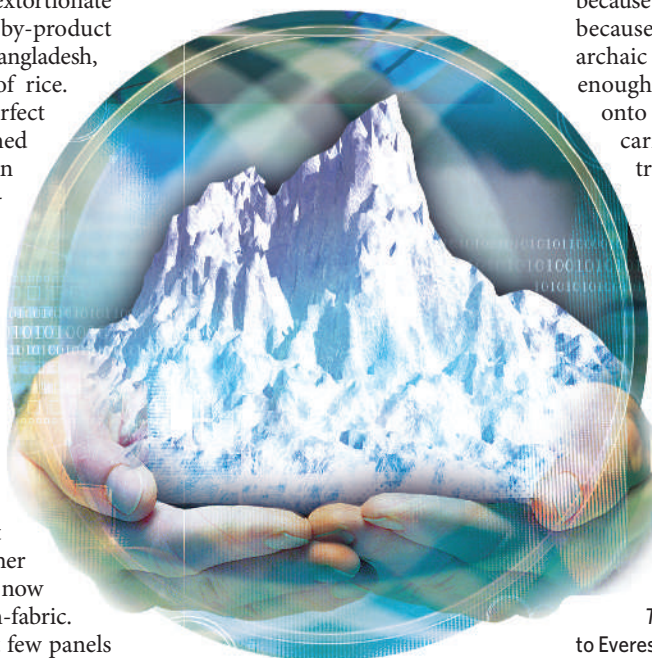
An hour later, she passed a group of purists. The night before, they'd refused to plug into any solar net other than the one they'd carried up themselves. That had restricted them to so little power that they'd had to use a dangerous-looking liquid-fuel stove rather than the microwave zappers favoured by everyone else. Courtney had been concerned enough to offer use of hers, but they'd refused.

Now, rather than wearing lightweight powersuits, they were bundled like giant puffballs — moving slowly, perhaps because of all that clothing, but more likely because they'd chosen some form of archaic equipment that didn't generate enough oxygen. Nor were they clipped onto the monofilament. Instead, they carried their own rope as they trudged slowly upward.

Whisking by them as if they were standing still, Courtney couldn't think of anything to say, so she contented herself with a polite nod. She'd never understood such people. It was just a hike. Why make it tougher than it had to be? ■

Richard A. Lovett is an award-winning science and science-fiction writer from Portland, Oregon. He has written for dozens of magazines including *New Scientist* and

The Economist. He has never been to Everest, but has skied in Greenland.



JACEY